Generalized convolution quadrature based on Runge-Kutta methods

Lopez-Fernandez, Maria ; Sauter, Stefan A

Abstract: In this paper, we develop the Runge-Kutta generalized convolution quadrature with variable time stepping for the numerical solution of convolution equations for time and space-time problems and present the corresponding stability and convergence analysis. For this purpose, some new theoretical tools such as tensorial divided differences, summation by parts with Runge-Kutta differences and a calculus for Runge-Kutta discretizations of generalized convolution operators such as an associativity property will be developed in this paper. Numerical examples will illustrate the stable and efficient behavior of the resulting discretization.

DOI: https://doi.org/10.1007/s00211-015-0761-2

Generalized Convolution Quadrature based on Runge-Kutta Methods

M. Lopez-Fernandez* S. Sauter†

July 18, 2014

Abstract

Convolution equations for time and space-time problems have many important applications, e.g., for the modelling of wave or heat propagation via ordinary and partial differential equations as well as for the corresponding integral equation formulations.

For their discretization, the *convolution quadrature* (CQ) has been developed since the late 1980's and is now one of the most popular method in this field.

However, the method and the theory are restricted to constant time stepping and only recently the *implicit Euler - generalized convolution quadrature* (gCQ) has been developed which allows for variable time stepping.

In this paper, we develop the gCQ for Runge-Kutta methods with variable time stepping and present the corresponding stability and convergence analysis. For this purpose, some new theoretical tools such as *tensorial divided differences*, summation by parts with *Runge-Kutta differences* and a calculus for Runge-Kutta discretizations of generalized convolution operators such as an *associativity property* will be developed in this paper.

Numerical examples will illustrate the stable and efficient behavior of the resulting discretization.

Keywords: wave equation, retarded potentials, boundary integral equations, convolution equations, convolution quadrature, variable step size, fast algorithms, contour integral methods.

Mathematics Subject Classification (2000): 65M15, 65R20, 65L06, 65M38

*Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, e-mail: `maria.lopez@math.uzh.ch`

†Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, e-mail: `stas@math.uzh.ch`

1 Introduction

Convolution operators play an important role in numerous applications which are modelled by linear time-invariant nonhomogeneous evolution equations. This includes problems in time and space-time wave and heat propagation problems which are formulated either by ordinary and partial differential equations or by the corresponding integral equations.

The discretization will be based on the *convolution quadrature* (CQ) method which has been developed originally by Lubich, see [12, 13, 16, 15] for parabolic problems and [14] for hyperbolic ones. The idea is to express the *convolution kernel* k as the inverse Laplace transform of some *transfer operator* K and to formulate the problem as an integro-differential equation in the Laplace domain.

The discretization then consists of approximating the (time-depending) differential equation in the Laplace domain by a time stepping method – besides multisteps methods also Runge-Kutta methods have been proposed and analyzed for this purpose [12, 13, 15, 3, 1, 2, 5]. The transformation back to the time domain results in a discrete convolution equation which then can be solved numerically. This method is nowadays one of the most popular method in this field.

However, the CQ method as well as its analysis relies strongly on the use of constant time stepping. In [11, 10], the *generalized convolution quadrature* (gCQ) has been introduced which allows for variable time stepping. The approach was limited to the first order implicit Euler scheme.

The goal of this paper is to introduce the *Runge-Kutta generalized convolution quadrature* which results in a method with much faster convergence rates as well as an improved long time behavior of the approximation compared to the implicit Euler method. The possibility to use variable time stepping allows to resolve adaptively a non-smooth behavior of the temporal solution which often occurs, e.g., in the short time range after an electric circuit is switched on and before it has reached a periodic state.

The paper is structured as follows. In Section 2 we will briefly recall the definition of one-sided convolution operators and define the class of convolution kernels which we will consider in this paper. In Section 3 we will introduce Runge-Kutta generalized convolution quadrature for the *discretization* of convolution operators. Its stability and convergence will be analyzed in Section 4 and the *summation-by-parts formula* for *divided Runge-Kutta differences* will be derived for this purpose. Section 5 is devoted to the numerical *solution* of convolution equations. We will present the discrete equations and derive an associativity property for the composition of Runge-Kutta generalized convolution operators which allows to use the stability and error analysis as in Section 4 to derive corresponding estimates for the discrete solution. Finally, we will report in Section 6 the results of numerical experiments to illustrate that, for problems where the regularity of the solution is not uniformly distributed in the time interval, our method converges with optimal convergence rates while other CQ-type methods are converging suboptimally.

2 The Class of Problems

We will consider the class of convolution operators as described in [14, Sec. 2.1] and recall its definition. Let B and D denote some normed vector spaces and let $\mathcal{L}(B, D)$ be the space of continuous, linear mappings. As a norm in $\mathcal{L}(B, D)$ we take the usual operator norm

$$\|\mathcal{F}\|_{D \leftarrow B} := \sup_{u \in B \setminus \{0\}} \frac{\|\mathcal{F}u\|_D}{\|u\|_B}.$$

For given $\phi : \mathbb{R}_{\geq 0} \to B$, we consider the convolution

$$\int_0^t k(t-\tau)\,\phi(\tau)\,d\tau \qquad \text{in } D \quad \text{for all } t \in [0, T]. \tag{1}$$

The kernel operator k is defined as the inverse Laplace transform of a given *transfer operator* K. The class of problems under consideration is defined as follows. For $\sigma \in \mathbb{R}$ we introduce

$$\mathbb{C}_\sigma = \{z \in \mathbb{C} \mid \operatorname{Re} z > \sigma\}.$$

Assumption 1 *For some $\sigma_K \in \mathbb{R}$ (describing the analyticity region) and some $\mu \in \mathbb{R}$ (describing the growth behavior), the class $\mathcal{A}^\mu_{\sigma_K}(B, D)$ of transfer operators consists of operator valued mappings $K : \mathbb{C}_{\sigma_K} \to \mathcal{L}(B, D)$ which satisfy:*

1. *$K : \mathbb{C}_{\sigma_K} \to \mathcal{L}(B, D)$ is analytic.*
2. *K satisfies the estimate*

$$\|K(z)\|_{D \leftarrow B} \leq C_{\mathrm{op}} (1 + |z|)^\mu, \qquad \forall z \in \mathbb{C}_{\sigma_K}, \tag{2}$$

for a fixed constant $C_{\mathrm{op}} > 0$.[1]

For $j \in \mathbb{Z}$, we define

$$K_j(z) := z^{-j} K(z). \tag{3}$$

For any

$$\nu \in \mathbb{N}_0 \quad \text{such that } \nu > \mu + 1, \tag{4}$$

the Laplace inversion formula

$$k_\nu(t) := \frac{1}{2\pi\,\mathrm{i}} \int_\gamma \mathrm{e}^{zt} K_\nu(z)\,dz, \tag{5}$$

for a contour $\gamma = \sigma + \mathrm{i}\,\mathbb{R}$, with $\sigma > \sigma_K$, defines a continuous and exponentially bounded operator $k_\nu(t)$, which by Cauchy's integral theorem vanishes for $t < 0$.

[1] The generic constant C in the following estimates will depend on C_{op} but not explicitly on σ_K. Hence, if C_{op} is independent of σ_K so is the constant C.

Let

$$C_0^j\left([0,T],B\right) := \left\{\psi \in C^j\left([0,T],B\right) \mid \forall 0 \le r \le j-1: \quad \psi^{(r)}(0) = 0\right\}.$$

As in [14] we denote the convolution $k * \phi$ for $\phi \in C_0^\nu\left([0,T],B\right)$ and ν as in (4) by

$$\left(K\left(\partial_t\right)\phi\right)(t) := \int_0^t k_\nu(\tau)\,\partial_t^\nu \phi\,(t-\tau)\,d\tau. \tag{6}$$

Then

$$\left(K\left(\partial_t\right)\phi\right)(t) = \int_0^t \left(\frac{1}{2\pi\,\mathrm{i}}\int_\gamma \mathrm{e}^{z\tau}\,K_\nu(z)dz\right)\partial_t^\nu\phi\,(t-\tau)\,d\tau, \tag{7}$$

where the integrals exist as Riemann integrals.

Remark 2 *Equation (7) can be rewritten as the coupled system*

$$\left(K\left(\partial_t\right)\phi\right)(t) = \frac{1}{2\pi\,\mathrm{i}}\int_\gamma (K_\nu(z)u_\nu\left(z,t\right)dz \tag{8a}$$

with the solution u_ν of

$$\partial_t u_\nu(z,t) = z u_\nu(z,t) + \partial_t^\nu\phi(t), \qquad u_\nu(z,0) = 0, \tag{8b}$$

and γ a suitable contour in the complex plane: either a vertical contour running from $\sigma - \mathrm{i}\,\infty$ to $\sigma + \mathrm{i}\,\infty$, for some ν which satisfies (4), or a suitable closed contour clockwise oriented.

3 Runge-Kutta Generalized Convolution Quadrature

3.1 Runge-Kutta Methods

The discretization of the convolution (6) will be based on a discretization of the ordinary differential equation by a Runge-Kutta method with variable time steps. In this section, we will introduce the class of Runge-Kutta methods which we will consider and collect some basic properties – for proofs and further details we refer to [8].

We consider Runge–Kutta method of s stages given by the Butcher table $\mathbf{A} = (a_{i,j})_{i,j=1}^s$, $\mathbf{b} = (b_i)_{i=1}^s$, $\mathbf{c} = (c_i)_{i=1}^s$. For the discretization we employ a sequence of time points $\Theta := (t_n)_{n=0}^N$ with

$$0 = t_0 < t_1 < \ldots < t_N = T, \quad \Delta_j = t_j - t_{j-1}, \quad \Delta := \max_{1\le i\le n}\Delta_j. \tag{9}$$

The local quasi-uniformity of the mesh is defined as the constant

$$c_\Theta := \frac{1}{2}\max_{2\le i\le N}\left(\frac{\Delta_i}{\Delta_{i-1}} + \frac{\Delta_{i-1}}{\Delta_i}\right). \tag{10a}$$

As a further (mild) assumption on the mesh width we impose the condition on the maximal mesh width

$$\Delta \leq C_\Theta / N. \tag{10b}$$

Notation 3 *The internal time points are defined by* $t_{n,i} = t_{n-1} + c_i \Delta_n$, $i = 1, \ldots, s$. *For a function* g *which is defined in the time interval* $[0, T]$, *we introduce*

$$\mathbf{g}^{(n)} := \left(g\left(t_{n,i}\right)\right)_{i=1}^{n} \in \mathbb{C}^s.$$

The time step n *is denoted as a superscript for vectors and matrices in order not to confuse with their components. The* m*-th time derivative of function* u *is denoted by* $\partial_t^m u$ *and its evaluation at some time point* t_k *is*

$$\partial_t^m u^{(k)} := \frac{d^m u}{dt^m}\left(t_k\right).$$

Further, we introduce $\mathbf{1} = (1)_{i=1}^{s}$ *and, for vectors* $\mathbf{v}, \mathbf{w} \in \mathbb{C}^s$, *the bilinear (not sesquilinear!) form*

$$\mathbf{v} \cdot \mathbf{w} := \sum_{j=1}^{s} v_j w_j.$$

We also recall here the Hadamard product of two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{C}^s$ *by*

$$\mathbf{v} \odot \mathbf{w} = \left(v_i w_i\right)_{i=1}^{s} \quad \textit{and} \quad \mathbf{v}^{m\odot} = \underbrace{\mathbf{v} \odot \ldots \odot \mathbf{v}}_{m\text{-}times}.$$

The application of the s-stage Runge-Kutta methods to the initial value problem $y' = f(t, y)$, $y(0) = y_0$ can be written as the following recursion

$$\begin{aligned} Y_i^{(n)} &= y^{(n-1)} + \sum_{j=1}^{s} a_{i,j} f\left(t_{n-1} + c_j \Delta_n, Y_j^{(n)}\right) \qquad i = 1, \ldots, s \\ y^{(n)} &= y^{(n-1)} + \sum_{j=1}^{s} b_j f\left(t_{n-1} + c_j \Delta_n, Y_j^{(n)}\right). \end{aligned}$$

The Runge-Kutta method has *(classical) order* $p \geq 1$ and *stage order* q if for sufficiently smooth right-hand side f

$$Y_i^{(1)} - y\left(c_i \Delta_1\right) = \mathcal{O}\left(\Delta_1^{q+1}\right) \quad \forall i = 1, \ldots, m \quad \text{and} \quad y^{(1)} - y\left(t_1\right) = \mathcal{O}\left(\Delta_1^{p+1}\right),$$

as $\Delta_1 \to 0$.

For the analysis of the Runge-Kutta method, the stability function

$$R(z) := 1 + z\mathbf{b} \cdot \left(\mathbf{I} - z\mathbf{A}\right)^{-1} \mathbf{1} \tag{11}$$

plays a central role; here, and in the following $\mathbf{I}$ denotes the identity matrix. Throughout the paper we assume that the Runge-Kutta method satisfies the following assumption.

Assumption 4
The Runge-Kutta method is A-stable, this is

$$|R(z)| \leq 1, \quad \textit{for } \operatorname{Re} z \leq 0, \tag{12}$$

with classical order $p \geq 1$ and stage order $q \leq p$ and it is stiffly accurate, this is

$$\mathbf{b} = \mathbf{A}^{\intercal}\mathbf{e}^{(s)} \quad \textit{with} \quad \mathbf{e}^{(s)} = (0, \ldots, 0, 1)^{\intercal} \in \mathbb{R}^s. \tag{13}$$

Remark 5 *In what follows we will repeatedly use the following properties of Runge–Kutta methods satisfying Assumption 4:*

1. *Condition (13) implies $R(\infty) = 0$ and $c_s = 1$ [8, Chap. IV, Prop. 3.8].*
2. *The assumption of A-stability implies that the coefficient matrix $\mathbf{A}$ is diagonalizable [8, Theorem 4.12] and all eigenvalues d_i, $1 \leq i \leq s$, have strictly positive real part. In particular $\mathbf{A}$ is invertible.*
3. *If the method has stage order q, it holds ([8, (15.5)])*

$$\mathbf{A}\mathbf{c}^{(m-1)\odot} = \frac{1}{m}\mathbf{c}^{m\odot} \qquad \forall 1 \leq m \leq q. \tag{14}$$

4. *If the method has order p, it follows (cf. [1, 16])*

$$\mathbf{b} \cdot \mathbf{A}^{\ell}\mathbf{c}^{(k-1)\odot} = \mathbf{b} \cdot \mathbf{A}^{\ell-1}\mathbf{c}^{k\odot}/k, \qquad \forall k + \ell \leq p. \tag{15}$$

3.2 Discretization of the Convolution Operator

The starting point of the discretization of the convolution operator is the representation (8). We will add more flexibility in the discretization by replacing the regularization parameter ν by a parameter $\rho \in \mathbb{N}_0$. The stability and convergence analysis will show that ρ can be chosen in the range

$$\nu - (q+1) \leq \rho \leq p + \nu - (q+1), \tag{16}$$

where $\nu > \mu + 1$ is as in (7), p is the order of the Runge–Kutta method which we will employ for the discretization and q is the stage order; some hints for the choice of ρ will be given in Remarks 7 and 18.

The discretization will be based on an approximation of the ordinary differential equation (cf. (8b))

$$\partial_t u_\rho(z,t) = z u_\rho(z,t) + \partial_t^\rho \phi(t), \qquad u_\rho(z,0) = 0.$$

Assumption 4 implies (13) so that the chosen Runge–Kutta method can be written in the form

$$\mathbf{u}_\rho^{(n)}(z) = \left(\mathbf{1} \otimes \mathbf{e}^{(s)}\right) \mathbf{u}_\rho^{(n-1)}(z) + \Delta_n \mathbf{A} \left(z\mathbf{u}_\rho^{(n)}(z) + \partial_t^\rho \boldsymbol{\phi}^{(n)}\right). \tag{17}$$

We can write (17) as a recurrence for $\mathbf{u}_\rho^{(n)}$

$$\begin{aligned}\mathbf{u}_\rho^{(n)}(z) &= (\mathbf{I}-\Delta_n z\mathbf{A})^{-1}\left(\left(\mathbf{1}\otimes\mathbf{e}^{(s)}\right)\mathbf{u}_\rho^{(n-1)}(z)+\Delta_n\mathbf{A}\partial_t^\rho\boldsymbol{\phi}^{(n)}\right) \qquad (18)\\ &= \left(\mathbf{R}(\Delta_n z)\otimes\mathbf{e}^{(s)}\right)\mathbf{u}_\rho^{(n-1)}(z)+\Delta_n(\mathbf{I}-z\Delta_n\mathbf{A})^{-1}\mathbf{A}\partial_t^\rho\boldsymbol{\phi}^{(n)}\end{aligned}$$

with

$$\mathbf{R}(z) := (\mathbf{I}-z\mathbf{A})^{-1}\mathbf{1}. \qquad (19)$$

From the identity

$$(\mathbf{I}-z\mathbf{A})^{-1}\mathbf{A} = \frac{1}{z}(\mathbf{I}-z\mathbf{A})^{-1}-\frac{1}{z}\mathbf{I} \qquad (20)$$

which holds for all square matrices $\mathbf{A}$ with regular resolvent, we conclude that that the last component $\mathbf{e}^{(s)}\cdot\mathbf{R}$ equals the stability function R (cf. (11)).

The last component in (18), $\left(\mathbf{u}_\rho^{(n)}\right)_s$ then defines the approximation of $u(t_n)$.

Definition 6 (Runge-Kutta Generalized Convolution Quadrature) *Let the transfer operator K satisfy (2) and let $\nu\in\mathbb{N}_0$ be the smallest integer such that $\nu>\mu+1$. Let $\phi\in C_0^\nu([0,T],B)$ and consider the convolution operation*

$$K(\partial_t)\phi(t) = \int_0^t\left(\frac{1}{2\pi\,\mathrm{i}}\int_\gamma \mathrm{e}^{z\tau}K_\nu(z)dz\right)\partial_t^\nu\phi(t-\tau)\,d\tau \qquad \forall t\in[0,T]. \qquad (21)$$

Let a Runge-Kutta method be given which satisfies Assumption 4. Then the discretization of (21) by Runge-Kutta Generalized Convolution Quadrature *is given by*

$$\left(K_\rho\left(\partial_t^\Theta\right)\partial_t^\rho\phi\right)^{(n)} := \frac{1}{2\pi\,\mathrm{i}}\int_\gamma K_\rho(z)\,\mathbf{u}_\rho^{(n)}(z)\,dz, \qquad n=1,2,\ldots \qquad (22)$$

with $\mathbf{u}_\rho^{(0)}=\mathbf{0}$ and

$$\mathbf{u}_\rho^{(n)}(z) = \left(\mathbf{R}(\Delta_n z)\otimes\mathbf{e}^{(s)}\right)\mathbf{u}_\rho^{(n-1)}(z)+\Delta_n(\mathbf{I}-z\Delta_n\mathbf{A})^{-1}\mathbf{A}\partial_t^\rho\boldsymbol{\phi}^{(n)}, \qquad n=1,2,\ldots.$$

The approximation of $K(\partial_t)\phi$ at time point t_n is given by the last component $\mathbf{e}^{(s)}\cdot\left(K_\rho\left(\partial_t^\Theta\right)\left(\times_{k=1}^N\boldsymbol{\phi}_\rho^{(k)}\right)\right)^{(n)}$. Here, $\rho\in\mathbb{N}_0$ is a regularization parameter which can be chosen in the range

$$\nu-(q+1)\le\rho\le p+\nu-(q+1),$$

where p is the classical order of the Runge–Kutta method and q denotes the stage order.

Remark 7 *It is important to mention that γ in (22), typically, is not chosen as the vertical contour $\sigma+\mathrm{i}\,\mathbb{R}$ but as a finite closed contour which encircles the poles of $\mathbf{u}_\rho^{(n)}$ and is contour clockwise oriented. For the practical realization the contour integral in (22) has to be approximated by numerical quadrature (see also Remark 18); for the implicit Euler method this has been developed and analyzed in [9, 10] while for Runge-Kutta method this is the topic of a forthcoming paper.*

4 Error Analysis of Runge-Kutta Generalized Convolution Quadrature

The analysis of the Runge-Kutta gCQ consists of several steps: First, we will resolve the recursion in (18) to express $\mathbf{u}_\rho^{(n)}$ as a sum over the history. This allows to employ a summation-by-parts formula which allows to gain negative powers of z (and hence a faster decay of the integrand for large z) on the expense of increased smoothness requirements on the input function ϕ.

4.1 Summation-by-Parts

The recursion (18) can be resolved and we obtain

$$\mathbf{u}_\rho^{(n)}(z) = \Delta_n \left(\mathbf{I} - z\Delta_n \mathbf{A}\right)^{-1} \mathbf{A}\partial_t^\rho \boldsymbol{\phi}^{(n)} + \sum_{k=1}^{n-1} \Delta_k \left(\prod_{\ell=k+1}^{n-1} R\left(\Delta_\ell z\right)\right) \left(\mathbf{e}^{(s)} \cdot \left(\mathbf{I} - z\Delta_k \mathbf{A}\right)^{-1} \mathbf{A}\partial_t^\rho \boldsymbol{\phi}^{(k)}\right) \mathbf{R}\left(\Delta_n z\right).$$

For the last component $\mathbf{e}^{(s)} \cdot \mathbf{u}_\rho^{(n)}(z)$ this formula simplifies and we obtain

$$\mathbf{e}^{(s)} \cdot \mathbf{u}_\rho^{(n)}(z) = \sum_{k=1}^{n} \Delta_k \left(\prod_{\ell=k+1}^{n} R\left(\Delta_\ell z\right)\right) \left(\mathbf{e}^{(s)} \cdot \left(\mathbf{I} - z\Delta_k \mathbf{A}\right)^{-1} \mathbf{A}\partial_t^\rho \boldsymbol{\phi}^{(k)}\right). \quad (23)$$

For the forthcoming analysis it is convenient to write this equation by using Kronecker matrices and tensor calculus. Let us then define the tensors

$$\mathbf{e}^{k\otimes} := \bigotimes_{\ell=1}^{k} \mathbf{e}^{(s)}, \qquad \mathbf{1}^{k\otimes} := \bigotimes_{\ell=1}^{k} \mathbf{1} \quad (24)$$

and the Kronecker matrix

$$\mathbb{A}^{(k,n)}(z) := \bigotimes_{\ell=k}^{n} \left(\mathbf{I} - z\Delta_\ell \mathbf{A}\right)^{-1}.$$

Recall that a Kronecker matrix $\bigotimes_{j=1}^{d} \mathbf{B}^{(j)}$ is applied to a tensor $\bigotimes_{j=1}^{d} \mathbf{v}^{(j)}$ of vectors $\mathbf{v}^{(j)}$ by means of

$$\left(\bigotimes_{j=1}^{d} \mathbf{B}^{(j)}\right) \left(\bigotimes_{j=1}^{d} \mathbf{v}^{(j)}\right) = \bigotimes_{j=1}^{d} \mathbf{B}^{(j)} \mathbf{v}^{(j)}.$$

The *canonical extension* of the bilinear form $\mathbf{v} \cdot \mathbf{w}$ to tensors is

$$\left(\bigotimes_{j=1}^{d} \mathbf{v}^{(j)}\right) \cdot \left(\bigotimes_{j=1}^{d} \mathbf{w}^{(j)}\right) = \prod_{j=1}^{d} \mathbf{v}^{(j)} \cdot \mathbf{w}^{(j)}.$$

Finally, the *vectorization is given by*

$$\left(\left(\bigotimes_{j=1}^{d-1}\mathbf{v}^{(j)}\right)\otimes\bullet\right)\cdot\left(\bigotimes_{j=1}^{d}\mathbf{w}^{(j)}\right):=\left(\left(\bigotimes_{j=1}^{d-1}\mathbf{v}^{(j)}\right)\cdot\left(\bigotimes_{j=1}^{d-1}\mathbf{w}^{(j)}\right)\right)\mathbf{w}^{(d)}$$
$$=\left(\prod_{j=1}^{d-1}\mathbf{v}^{(j)}\cdot\mathbf{w}^{(j)}\right)\mathbf{w}^{(d)}.$$

Then, we have

$$\mathbf{u}_{\rho}^{(n)}(z)=\sum_{k=1}^{n}\Delta_k\left(\mathbf{e}^{(n-k)\otimes}\otimes\bullet\right)\cdot\left(\mathbb{A}^{(k,n)}(z)\left(\mathbf{A}\partial_t^{\rho}\boldsymbol{\phi}^{(k)}\otimes\mathbf{1}^{(n-k)\otimes}\right)\right). \quad (25)$$

In the next step, we will introduce difference operators which are related to the time steps t_k and we will discuss their relation to Newton's divided differences later. Let again $\Theta:=(t_n)_{n=1}^{N}$ denote the time grid with steps $\Delta_j=t_j-t_{j-1}$. Formally we extend the time grid to the negative time axes by setting $t_{-j}=-j\Delta_1$, $j\in\mathbb{N}$.

Definition 8 (Divided Runge-Kutta Differences) *Let a Runge-Kutta method be given by the Butcher table* $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ *with non-singular* $\mathbf{A}$. *For a subset* $\mathcal{I}\subset\mathbb{Z}$ *of consecutive integers, let* $\Theta_{\mathcal{I}}:=(x_k)_{k\in\mathcal{I}}\subset\mathbb{R}$ *denote a sequence of strictly increasing points with steps* $\Delta_k=x_k-x_{k-1}$. *We set*

$$\mathcal{I}'=\{k\in\mathbb{Z}\mid\{k-1,k\}\subset\mathcal{I}\}.$$

For a function v *which is defined in the points* $x_{k,r}:=x_{k-1}+c_r\Delta_k$, *for all* $k\in\mathcal{I}'$ *and* $1\le r\le s$, *the* Runge-Kutta differences $[\![\ldots]\!]v$ *are given by the recursion:*

$$[\![x_k]\!]v:=\mathbf{v}^{(k)}:=\left(v\left(x_{k,r}\right)\right)_{r=1}^{s}\qquad\forall k\in\mathcal{I}' \quad (26)$$

and for all $i,k\in\mathcal{I}'$ *with* $i<k$

$$[\![x_i,x_{i+1},\ldots,x_k]\!]v:=\left(\Delta_k\mathbf{A}\right)^{-1}\left([\![x_{i+1},\ldots,x_k]\!]v-\left(\mathbf{1}\otimes\mathbf{e}^{(s)}\right)[\![x_i,\ldots,x_{k-1}]\!]v\right). \quad (27)$$

For $m\in\mathbb{N}_0$, *the tuple of* m-th order Runge-Kutta differences $[\![\Theta_{\mathcal{I}}]\!]^m w\in\bigtimes_{k\in\mathcal{I}}\mathbb{C}^s$ *is given by*

$$[\![\Theta_{\mathcal{I}}]\!]^m v:=\bigtimes_{k\in\mathcal{I}}[\![x_{k-m},\ldots,x_k]\!]v. \quad (28)$$

For a tuple $\mathbb{V}=\bigtimes_{j\in\mathcal{I}'}\mathbf{v}^{(j)}$ *of vectors* $\mathbf{v}^{(j)}=\left(v_m^{(j)}\right)_{m=1}^{s}\in\mathbb{C}^s$ *we set*

$$[\![x_i,\ldots,x_k]\!]\mathbb{V}:=[\![x_i,\ldots,x_k]\!]v$$

for any continuous function v *which interpolates* $\mathbb{V}$ *at the mesh points, i.e.,* $v\left(x_{k,r}\right)=v_r^{(k)}$ *for all* $k\in\mathcal{I}'$ *and* $1\le r\le s$.

In particular we have (cf. (26))

$$[\![x_{k-1}, x_k]\!] v = (\Delta_k \mathbf{A})^{-1} \left(\mathbf{v}^{(k)} - \left(\mathbf{1} \otimes \mathbf{e}^{(s)} \right) \mathbf{v}^{(k-1)} \right). \tag{29}$$

Proposition 9 (Summation by parts formula) *Let a Runge-Kutta method be given by the Butcher table* $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ *with non-singular matrix* $\mathbf{A}$. *Let* $w : \mathbb{R}_{\geq 0} \to \mathbb{C}$ *be a function which can be continuously extended to* $\mathbb{R}_{<0}$ *by zero. The time mesh satisfies (9) and is extended by* $t_{-j} = -j\Delta_1$ *for* $j \in \mathbb{N}$. *Set* $\mathbf{w}^{(j)} = \left(w\left(t_{j,r}\right)\right)_{r=1}^{s} \in \mathbb{C}^s$, $j \in \mathbb{Z}_{\leq N}$ *and let* $\mathbf{e}^{r\otimes}$, $\mathbf{1}^{r\otimes}$ *be as in (24). Then, for any* $m \in \mathbb{N}_0$

$$\begin{aligned} &\sum_{k=0}^{n} \Delta_k \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet \right) \cdot \left(\mathbb{A}^{(k,n)}(z) \left(\mathbf{A}\mathbf{w}^{(k)} \otimes \mathbf{1}^{(n-k)\otimes} \right) \right) \\ &\quad = -\sum_{\ell=0}^{m-1} \frac{[\![t_{n-\ell}, \ldots, t_n]\!] w}{z^{\ell+1}} \\ &\quad + \frac{1}{z^m} \sum_{k=0}^{n} \Delta_k \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet \right) \cdot \left(\mathbb{A}^{(k,n)}(z) \left(\mathbf{A}[\![t_{k-m}, \ldots, t_k]\!] w \otimes \mathbf{1}^{(n-k)\otimes} \right) \right). \end{aligned} \tag{30}$$

For the corresponding generalized discrete convolution operator it holds

$$V\left(\partial_t^{\Theta}\right) w = V_m\left(\partial_t^{\Theta}\right) [\![\Theta]\!]^m w. \tag{31}$$

Proof. We denote the left-hand side in (30) by lhs and obtain (cf. (20))

$$\begin{aligned}
\mathrm{lhs} &= \sum_{k=0}^{n} \Delta_k \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \\
&\qquad \cdot \left(\left(\mathbf{I} - z\Delta_k \mathbf{A}\right)^{-1} \mathbf{A} \otimes \mathbb{A}^{(k+1,n)}(z)\right) \left(\mathbf{w}^{(k)} \otimes \mathbf{1}^{(n-k)\otimes}\right) \\
&\overset{(20)}{=} \frac{1}{z} \sum_{k=0}^{n} \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \\
&\qquad \cdot \left(\left(\left(\mathbf{I} - z\Delta_k \mathbf{A}\right)^{-1} - \mathbf{I}\right) \otimes \mathbb{A}^{(k+1,n)}(z)\right) \left(\mathbf{w}^{(k)} \otimes \mathbf{1}^{(n-k)\otimes}\right) \\
&= -\frac{\mathbf{w}^{(n)}}{z} + \frac{1}{z}\left(\mathbf{I} - z\Delta_k \mathbf{A}\right)^{-1} \mathbf{w}^{(n)} \\
&\quad + \frac{1}{z} \sum_{k=0}^{n-1} \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \\
&\qquad \cdot \left(\left(\mathbf{I} - z\Delta_k \mathbf{A}\right)^{-1} \otimes \mathbb{A}^{(k+1,n)}(z)\right) \left(\mathbf{w}^{(k)} \otimes \mathbf{1}^{(n-k)\otimes}\right) \\
&\quad - \frac{1}{z} \sum_{k=0}^{n-1} \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \left(\mathbf{I} \otimes \mathbb{A}^{(k+1,n)}(z)\right) \left(\mathbf{w}^{(k)} \otimes \mathbf{1}^{(n-k)\otimes}\right) \\
&= -\frac{\mathbf{w}^{(n)}}{z} + \frac{1}{z} \sum_{k=0}^{n} \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \mathbb{A}^{(k,n)}(z) \left(\mathbf{w}^{(k)} \otimes \mathbf{1}^{(n-k)\otimes}\right) \\
&\quad - \frac{1}{z} \sum_{k=1}^{n} \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \mathbb{A}^{(k,n)}(z) \left(\left(\mathbf{1} \otimes \mathbf{e}^{(s)}\right) \mathbf{w}^{(k-1)} \otimes \mathbf{1}^{(n-k)\otimes}\right) \\
&= -\frac{\mathbf{w}^{(n)}}{z} + \frac{1}{z} \sum_{k=0}^{n} \Delta_k \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \mathbb{A}^{(k,n)}(z) \left(\mathbf{A}[\![t_{k-1}, t_k]\!] w \otimes \mathbf{1}^{(n-k)\otimes}\right).
\end{aligned}$$

This one-fold summation by parts can be iterated and leads to the assertion.

The second relation (31) is a simple consequence of Cauchy's integral theorem. ■

The following proposition states the boundedness of the right-hand side in (30) with respect to a decreasing step size in terms of the stage order of the underlying Runge–Kutta method.

Definition 10 *Let $r \in \mathbb{N}_0$, $T > 0$, and V be a normed vector space with norm $\|\cdot\|_V$. For a vector-valued function $\mathbf{v} \in V^s$, we set*

$$\|\mathbf{v}\|_V := \max_{1 \le i \le s} \|v_i\|_V$$

if no confusion is possible.

For a function $w \in C^r([0,T], V)$ and any interval $\tau \subset [0,T]$, we set

$$|w|_{C^r(\tau,V)} := \frac{1}{r!} \sup_{t \in \tau} \|\partial^r w(t)\|_V \quad \text{and} \quad \|w\|_{C^r(\tau,V)} := \max_{0 \le \ell \le r} |v|_{C^\ell(\tau,V)}.$$

Proposition 11 *Let a Runge-Kutta method be given by the Butcher table* $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ *with non-singular* $\mathbf{A}$. *Let* V *be a normed vector space. If the method has stage order* q *then for* $0 \le \ell \le q+1$ *and any* $w \in C^{q+1}\left([t_{k-\ell}, t_k], V\right)$ *it holds*

$$\begin{aligned}[\![t_{k-\ell}, t_{k+1-\ell}, \dots, t_k]\!] w &= \partial_t^\ell \mathbf{w}^{(k)} + \mathbf{T}^{(k)}_{q+1-\ell}, \\ \left\| \mathbf{T}^{(k)}_{q+1-\ell} \right\|_V &\le C \left| w \right|_{C^{q+1}([t_{k-\ell}, t_k], V)} \Delta_k^{q+1-\ell},\end{aligned}$$

where C *depends on* c_Θ *(cf. (10a)),* q, *and* $\mathbf{A}$.

Proof. The proof is by induction. For $\ell = 0$ the result is obvious and we even have equality: $[\![t_k]\!] w = \mathbf{w}^{(k)}$ so that $\mathbf{T}^{(k)}_{q+1} = \mathbf{0}$.

Let us assume now that the result is true for $\ell - 1$. Then for ℓ we have

$$[\![t_{k-\ell}, t_{k-\ell+1}, \dots, t_k]\!] w = \Delta_k^{-1} \mathbf{A}^{-1} \left([\![t_{k-\ell+1}, \dots, t_k]\!] w - \left(\mathbf{1} \otimes \mathbf{e}^{(s)} \right) [\![t_{k-\ell}, \dots, t_{k-1}]\!] w \right) \tag{32}$$

$$\begin{aligned}&= \Delta_k^{-1} \mathbf{A}^{-1} \left(\partial_t^{\ell-1} \mathbf{w}^{(k)} - \left(\mathbf{1} \otimes \mathbf{e}^{(s)} \right) \partial_t^{\ell-1} \mathbf{w}^{(k-1)} + \tilde{\mathbf{T}}^{(k)}_{q+1-\ell} \right) \\ &= \Delta_k^{-1} \mathbf{A}^{-1} \left(\left(\int_{t_{k-1}}^{t_{k,m}} \partial_t^\ell w \right)_{m=1}^{s} + \tilde{\mathbf{T}}^{(k)}_{q+1-\ell} \right),\end{aligned}$$

where

$$\tilde{\mathbf{T}}^{(k)}_{q+1-\ell} := \mathbf{T}^{(k)}_{q+2-\ell} - \left(\mathbf{1} \otimes \mathbf{e}^{(s)} \right) \mathbf{T}^{(k-1)}_{q+2-\ell}.$$

Conditions (14) imply

$$\Delta_j \mathbf{A} \partial_t^\ell \mathbf{w}^{(j)} = \left(\int_{t_{j-1}}^{t_{j,m}} \partial_t^\ell w \right)_{m=1}^{s} + \boldsymbol{\xi}^{(j)},$$

with

$$\left\| \boldsymbol{\xi}^{(j)} \right\|_V \le C_q \Delta_j^{r+1} \left\| \partial_t^{\ell+r} w \right\|_{C^0(\tau_j, V)} \qquad 0 \le r \le q.$$

We apply this for $r = q + 1 - \ell$ and obtain

$$\Delta_j^{-1} \mathbf{A}^{-1} \left(\int_{t_{j-1}}^{t_{j,m}} \partial_t^\ell w \right)_{m=1}^{s} = \partial_t^\ell \mathbf{w}^{(j)} + \widetilde{\boldsymbol{\xi}}^{(j)} \tag{33}$$

with

$$\left\| \widetilde{\boldsymbol{\xi}}^{(j)} \right\|_V \le C_q C_{\mathbf{A}} \left| w \right|_{C^{q+1}([t_{j-1}, t_j], V)} \Delta_j^{q+1-\ell}. \tag{34}$$

The combination of (32) with the induction hypothesis, (33), and (34) yields

$$[\![t_{k-\ell}, t_{k-\ell+1}, \dots, t_k]\!] w = \partial_t^\ell \mathbf{w}^{(j)} + \mathbf{T}^{(k)}_{q+1-\ell}$$

with

$$\left\| \mathbf{T}^{(k)}_{q+1-\ell} \right\|_V \le C \left| w \right|_{C^{q+1}([t_{j-1}, t_j], V)} \Delta_j^{q+1-\ell}$$

and the result follows. ■

4.2 Stability

The starting point of the error estimates for the Runge-Kutta gCQ is the summation formula with summation by parts (cf. (31)):

$$K_\rho\left(\partial_t^\Theta\right)\partial_t^\rho\phi = K_{\rho+m}\left(\partial_t^\Theta\right)[\![\Theta]\!]^m\partial_t^\rho\phi. \tag{35}$$

Note that the A-stability assumption in (12) implies in particular that all poles of R have positive real part. These poles are given by $z = 1/d_i$ with the eigenvalues d_i of $\mathbf{A}$. This property allows to derive the following estimates.

Lemma 12 *Let the Runge-Kutta method be A-stable. Let d_i, $i = 1, \dots, s$, be the eigenvalues of the coefficient matrix $\mathbf{A}$. We set*

$$r_0 = \min\left\{\frac{\operatorname{Re} d_i}{|d_i|^2} : 1 \le i \le s\right\} > 0 \quad \textit{and} \quad \alpha_0 = \min\{|d_i| : 1 \le i \le s\} > 0. \tag{36}$$

(i) There exists a constant C depending on r_0 and the Runge-Kutta coefficients such that

$$|R(z)| \le 1 + C(\operatorname{Re} z)_+ \qquad \forall z \in \mathbb{C} \textit{ with } \operatorname{Re} z \le \frac{r_0}{2} \tag{37}$$

and $(x)_+ := \max\{0, x\}$.

(ii) Let $\mathbf{A} = \mathbf{V}^{-1}\mathbf{D}\mathbf{V}$ (cf. Remark 2). Then, it holds

$$\left\|(\mathbf{I} - z\mathbf{A})^{-1}\right\| \le \beta_0 := \frac{2}{\alpha_0 r_0}\left\|\mathbf{V}^{-1}\right\|\|\mathbf{V}\| \qquad \forall z \in \mathbb{C} \textit{ with } \operatorname{Re} z \le \frac{r_0}{2}. \tag{38}$$

Proof. (i) By using $\operatorname{Re}\left(\frac{1}{\zeta}\right) = (\operatorname{Re}\zeta)/|\zeta|^2$, we conclude that R is analytic for all $z \in \mathbb{C}$ with $\operatorname{Re} z < r_0$. Then there exists $C_R > 0$ such that $|R(z)| \le C_R$ for all $z \in \mathbb{C}$ with $\operatorname{Re} z \le \frac{3}{4}r_0$. We conclude from Cauchy's integral theorem that $|R'(z)| \le \frac{4C_R}{r_0}$ for all $z \in \mathbb{C}$ with $\operatorname{Re} z \le \frac{r_0}{2}$. Taylor's theorem gives us the estimate

$$|R(x + \mathrm{i}\, y)| \le |R(\mathrm{i}\, y)| + \frac{4C_R}{r_0}x \qquad \forall 0 \le x \le r_0/2 \text{ and } y \in \mathbb{R}.$$

Since A-stability implies $|R(\mathrm{i}\, y)| \le 1$ we conclude that

$$|R(z)| \le 1 + C\operatorname{Re} z \qquad \forall z \in \mathbb{C} \text{ with } 0 \le \operatorname{Re} z \le r_0/2$$

holds. Estimate (37) is trivial for $\operatorname{Re} z \le 0$ (cf. (12))

(ii) By Remark 2 we can estimate

$$\left\|(\mathbf{I} - z\mathbf{A})^{-1}\right\| \le \left\|\mathbf{V}^{-1}\right\|\|\mathbf{V}\| \max_{1\le i\le s}\left\{\frac{1}{|1 - zd_i|}\right\}.$$

Writing $z = x + \mathrm{i}\, y$ and $d_i = u + \mathrm{i}\, v$, we obtain

$$|1 - z d_i|^2 = (1 - xu + yv)^2 + (yu + xv)^2 =: \kappa(y).$$

The quadratic function κ attains its minimum at $y = -\frac{v}{u^2+v^2}$ so that

$$\kappa(y) \geq \frac{\left(u - x\left(u^2+v^2\right)\right)^2}{u^2+v^2}.$$

Note that for $0 \leq x \leq \frac{u}{2(u^2+v^2)}$, it holds

$$\kappa(y) \geq \frac{\left(x\left(u^2+v^2\right) - u\right)^2}{u^2+v^2} \geq \frac{1}{4}\frac{u^2}{u^2+v^2}.$$

This proves (38). ■

Theorem 13 *Let a Runge-Kutta method be given by the Butcher table* $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$, *has stage order* q, *and satisfy Assumption 4. Fix* $\sigma_0 \geq \sigma_K$ *and let the maximal step* Δ *satisfy*

$$\frac{r_0}{2} - \Delta\sigma_0 \geq 0. \tag{39}$$

Let $\tilde{\rho} \in \mathbb{N}_0$ *be such that* $\nu - (q+1) \leq \tilde{\rho} \leq \nu$ *holds. Assume that* $\phi \in C_0^{\tilde{\rho}}([0,T], D)$. *Then, for any* $\tilde{m} \in \mathbb{N}_0$ *with*

$$\mu - \tilde{\rho} + 1 < \tilde{m} \leq q + 1, \tag{40}$$

the stability estimate

$$\left\| \left(K_{\tilde{\rho}}\left(\partial_t^{\Theta}\right) \partial_t^{\tilde{\rho}} \phi \right)^{(n)} \right\|_D \leq C \sum_{k=0}^{n} \Delta_k \, \mathrm{e}^{C\sigma_0(t_n - t_k)} \left\| [\![t_{k-\tilde{m}}, \ldots, t_k]\!] \partial_t^{\tilde{\rho}} \phi \right\|_B \tag{41}$$

holds. If $\phi \in C^{\tilde{\rho}+\tilde{m}}([0,T], D)$ *then*

$$\left\| K_{\tilde{\rho}}\left(\partial_t^{\Theta}\right) \partial_t^{\tilde{\rho}} \phi \right\|_D \leq C\, \mathrm{e}^{C\sigma_0 T} \left\| \partial_t^{\tilde{\rho}+\tilde{m}} \phi \right\|_{C^0([0,T],B)}. \tag{42}$$

Proof. By Proposition 11 the $\tilde{m}$-th order divided Runge-Kutta difference of $\partial_t^{\tilde{\rho}}\phi$ are bounded and we apply $\tilde{m}$-times summation by parts, i.e., consider (35) for $\tilde{m}$ as in (40). The assumption (40) ensures that the contour in the definition of the generalized convolution $K_{\tilde{\rho}+\tilde{m}}\left(\partial_t^{\Theta}\right)$ can be chosen as the vertical axes $\gamma = \sigma + \mathrm{i}\,\mathbb{R}$. Note that (35) equals

$$\left(K_{\tilde{\rho}}\left(\partial_t^{\Theta}\right) \partial_t^{\tilde{\rho}} \phi \right)^{(n)} = \frac{\Delta_n}{2\pi\,\mathrm{i}} \int_\gamma K_{\tilde{\rho}+\tilde{m}}(z) \left(z\Delta_n \mathbf{I} - \mathbf{A}^{-1} \right)^{-1} dz \left([\![t_{n-\tilde{m}}, \ldots, t_n]\!] \partial_t^{\tilde{\rho}} \phi \right) \tag{43}$$

$$+ \sum_{k=0}^{n-1} \frac{\Delta_k}{2\pi\,\mathrm{i}} \int_\gamma K_{\tilde{\rho}+\tilde{m}}(z) \left(\mathbf{I} - z\Delta_n \mathbf{A}^{-1} \right)^{-1} \mathbf{1}$$

$$\cdot \left(\mathbf{e}^{(s)} \cdot \left(z\Delta_k \mathbf{I} - \mathbf{A}^{-1} \right)^{-1} [\![t_{k-\tilde{m}}, \ldots, t_k]\!] \partial_t^{\tilde{\rho}} \phi \right) \prod_{\ell=k+1}^{n-1} \left(\mathbf{e}^{(s)} \cdot (\mathbf{I} - z\Delta_\ell \mathbf{A})^{-1} \mathbf{1} \right) dz.$$

Assumption (13) implies that

$$R(z) = \mathbf{e}^{(s)} \cdot (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{1}$$

and then by Lemma 12 we can bound

$$\left| \prod_{\ell=k+1}^{n-1} \left(\mathbf{e}^{(s)} \cdot (\mathbf{I} - z\Delta_\ell \mathbf{A})^{-1} \mathbf{1} \right) \right| \leq \prod_{\ell=k+1}^{n-1} (1 + C\sigma_0 \Delta_\ell) \leq \mathrm{e}^{C\sigma_0 (t_{n-1} - t_k)}.$$

Furthermore, we have

$$\left\| \mathbf{e}^{(s)} \cdot \left(z\Delta_k \mathbf{I} - \mathbf{A}^{-1} \right)^{-1} [\![t_{k-\tilde{m}}, \ldots, t_k]\!] \partial_t^{\tilde{\rho}} \phi \right\|_B \leq \beta_0 \|\mathbf{A}\| \left\| [\![t_{k-\tilde{m}}, \ldots, t_k]\!] \partial_t^{\tilde{\rho}} \phi \right\|_B$$

and

$$\left\| \left(z\Delta_n \mathbf{I} - \mathbf{A}^{\cdot^{-1}} \right)^{-1} \left([\![t_{k-\tilde{m}}, \ldots, t_k]\!] \partial_t^{\tilde{\rho}} \phi \right) \right\|_B \leq \beta_0 \|\mathbf{A}\| \left\| [\![t_{k-\tilde{m}}, \ldots, t_k]\!] \partial_t^{\tilde{\rho}} \phi \right\|_B.$$

Hence,

$$\begin{aligned} &\left\| \left(K_{\tilde{\rho}} \left(\partial_t^{\Theta} \right) \partial_t^{\tilde{\rho}} \phi \right)^{(n)} \right\|_D \qquad (44) \\ &\quad \leq \sqrt{s} \frac{(\beta_0 \|\mathbf{A}\|)^2}{2\pi} \sum_{k=0}^{n} \Delta_k \, \mathrm{e}^{C\sigma_0 (t_n - t_k)} \left\| [\![t_{k-\tilde{m}}, \ldots, t_k]\!] \partial_t^{\tilde{\rho}} \phi \right\|_D \int_\gamma |z|^{\mu - \tilde{\rho} - \tilde{m}} \, dz \end{aligned}$$

with an adjusted value of β_0. The choice of $\tilde{\rho}$ as stated in the lemma implies

$$\left\| \left(K_{\tilde{\rho}} \left(\partial_t^{\Theta} \right) \partial_t^{\tilde{\rho}} \phi \right)^{(n)} \right\|_D \leq C \sum_{k=0}^{n} \Delta_k \, \mathrm{e}^{C\sigma_0 (t_n - t_k)} \left\| [\![t_{k-\tilde{m}}, \ldots, t_k]\!] \partial_t^{\tilde{\rho}} \phi \right\|_B \qquad (45)$$

with $C := \sqrt{s} \frac{(\beta_0 \|\mathbf{A}\|)^2}{2\pi} \int_\gamma |z|^{\mu - \tilde{\rho} - \tilde{m}} \, dz$, which is (41). The combination with Proposition 11 gives (42). ■

4.3 Convergence

Theorem 14 *Let $K \in \mathcal{A}^{\mu}_{\sigma_K}(B, D)$ be a transfer operator and let $\nu \in \mathbb{N}_0$ denote the smallest integer with $\nu > \mu + 1$. Let an A-stable Runge-Kutta method be given by the Butcher table $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$, have stage order $q \geq 1$, order $p \geq q + 1$ and satisfy Assumption 4. Fix $\sigma \geq \sigma_K$ and let the maximal step Δ (cf. (9))satisfy*

$$\frac{r_0}{2} - \Delta\sigma \geq 0, \qquad (46)$$

with r_0 in (36).

For any $\rho \in \mathbb{N}_{\geq 0}$ in (22) with $\rho \geq \nu - (q+1)$ and $\phi \in C_0^{\nu}([0,T], B)$ let

$$w := K(\partial_t)\, \phi \quad \text{and} \quad w_\rho^{(n)} := \mathbf{e}^{(s)} \cdot \left(K_\rho \left(\partial_t^{\Theta} \right) \partial_t^{\rho} \phi \right)^{(n)}.$$

Then, the error estimate

$$\left\| w\left(t_n\right) - w_\rho^{(n)} \right\|_D \leq C \begin{cases} \|\phi\|_{C^{\rho+p+1}([0,T],B)}\, c_{\mu-\rho+p-q-1}\left(\Delta\right) \Delta^{\min\{p,\rho+q-\mu\}} & \mu-\rho<-1, \\ \|\phi\|_{C^{\nu+p+1}([0,T],B)}\, \Delta^{\min\{p,\rho+q+1-\nu\}} & \mu-\rho\geq -1 \end{cases} \tag{47}$$

holds with

$$c_\nu\left(\Delta\right) := \begin{cases} 1 & \nu \neq -1, \\ \log \frac{1}{\Delta} & \nu = -1 \end{cases}$$

provided that $\phi^{(r)}(0) = 0$ *for all* $r = 0, \ldots, \rho+q$ *and* $\phi \in C^{\nu+p+1}\left([0,T], B\right)$.

Note that estimate (47) implies that the choice $\rho = p+\nu-(q+1)$ (cf. (16)) leads to a convergence order $\mathcal{O}\left(\Delta^p\right)$ for sufficiently smooth and compatible data; for a further discussion see Remarks 7 and 18.

Proof. We assume in more generality that

$$\phi^{(r)}(0) = 0 \qquad \forall r = 0, \ldots, \rho+m-1$$

and choose $m \leq q+1$ later in an appropriate way.

Further we introduce the solution of the Runge-Kutta gCQ with right-hand side $[\![\Theta]\!]^m \partial_t^\rho \phi$, given by (see (28) and (25))

$$\mathbf{u}_{\rho,m}^{(n)}\left(z\right) = \sum_{k=1}^{n} \Delta_k \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet \right) \cdot \left(\mathbb{A}^{(k,n)}\left(z\right) \left(\mathbf{A} [\![t_{k-m}, \ldots, t_k]\!] \partial_t^\rho \phi \otimes \mathbf{1}^{(n-k)\otimes} \right) \right). \tag{48}$$

As usual, the last component is denoted by $u_{\rho,m}^{(n)} := \mathbf{e}^{(s)} \cdot \mathbf{u}_{\rho,m}^{(n)}\left(z\right)$.

Case 1: $\mu - \rho < -1$.

In this case (8a) and (22) hold for any $\rho \geq 0$ and we have

$$\delta w^{(n)} := w(t_n) - w_\rho^{(n)} = \frac{1}{2\pi\,\mathrm{i}} \int_\gamma K_\rho\left(z\right) \left(u_\rho\left(z, t_n\right) - u_\rho^{(n)}\left(z\right) \right) dz. \tag{49}$$

We choose the contour $\gamma = \sigma + \mathrm{i}\,\mathbb{R}$ and split it into

$$\gamma_{\text{near}} := \{ \zeta \in \gamma : |\zeta\Delta| < C_{\text{split}} \} \quad \text{and} \quad \gamma_{\text{far}} := \gamma \backslash \gamma_{\text{near}} \tag{50}$$

with some $0 < C_{\text{split}} = \mathcal{O}\left(1\right)$ which will be fixed later. This induces the splitting

$$\delta w_{\text{near}}^{(n)} := \frac{1}{2\pi\,\mathrm{i}} \int_{\gamma_{\text{near}}} K_\rho\left(z\right) \left(u_\rho\left(z, t_n\right) - u_\rho^{(n)}\left(z\right) \right) dz \quad \text{and} \quad \delta w_{\text{far}}^{(n)} := \delta w^{(n)} - \delta w_{\text{near}}^{(n)}.$$

Far Field

For the farfield estimates, we restrict to $m \leq q+1$. In order to estimate the component of (49) which is related to the farfield we will estimate the difference

$u_\rho(z,t_n) - u_\rho^{(n)}(z)$ for $z \in \gamma_{\text{far}}$. On the one side we observe that the exact solution of the ODE is given by

$$u_\rho(z,t) = \int_0^t \mathrm{e}^{z(t-\tau)}\, \partial_t^\rho \phi(\tau)\, d\tau. \tag{51}$$

Since $\partial_t^{\rho+\ell}\phi(0) = 0$ for $0 \le \ell \le m-1 \le q$ and $\phi \in C^{\rho+m}([0,T])$, we get via partial integration

$$u_\rho(z,t) = -\sum_{\ell=0}^{m-1} \frac{\partial_t^{\rho+\ell}\phi(t)}{z^{\ell+1}} + \frac{u_{\rho+m}(z,t)}{z^m}. \tag{52}$$

On the other side, we recall that the numerical approximation by the Runge–Kutta method can be written by using tensor notation as in (25), this is

$$\mathbf{u}_\rho^{(n)}(z) = \sum_{k=1}^{n} \Delta_k \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet \right) \cdot \left(\mathbb{A}^{(k,n)}(z) \left(\mathbf{A} \partial_t^\rho \boldsymbol{\phi}^{(k)} \otimes \mathbf{1}^{(n-k)\otimes} \right) \right).$$

Summation by parts (Proposition 9) yields

$$\mathbf{u}_\rho^{(n)}(z) = -\sum_{\ell=0}^{m-1} \frac{[\![t_{n-\ell},\dots,t_n]\!]\partial_t^\rho \phi}{z^{\ell+1}} + \frac{\mathbf{u}_{\rho,m}^{(n)}(z)}{z^m}, \tag{53}$$

with $\mathbf{u}_{\rho,m}^{(n)}$ as in (48). Since $u_\rho^{(n)} = \mathbf{e}^{(s)} \cdot \mathbf{u}_\rho^{(n)}$ the error can be written in the form

$$\delta w_{\text{far}}^{(n)} = \sum_{\ell=0}^{m-1} \delta w_{\text{far},\ell}^{(n)} + w_{\text{far},\rho,m}^{(n)} - w_{\text{far},m}(t_n) \tag{54}$$

with

$$\begin{aligned}
\delta w_{\text{far},\ell}^{(n)} &:= \frac{1}{2\pi\,\mathrm{i}} \int_{\gamma_{\text{far}}} \frac{K_\rho(z)}{z^{\ell+1}} \left(\mathbf{e}^{(s)} \cdot [\![t_{n-\ell},\dots,t_n]\!]\partial_t^\rho\phi - \partial_t^{\rho+\ell}\phi(t_n) \right) dz, \\
w_{\text{far},m}^{(n)} &:= \frac{1}{2\pi\,\mathrm{i}} \int_{\gamma_{\text{far}}} \frac{K_\rho(z)}{z^m} u_{\rho,m}^{(n)}(z)\, dz, \\
w_{\text{far},m}(t_n) &:= \frac{1}{2\pi\,\mathrm{i}} \int_{\gamma_{\text{far}}} \frac{K_\rho(z)}{z^m} u_{\rho+m}(z,t_n)\, dz.
\end{aligned}$$

Proposition 11 implies

$$\left\| \mathbf{e}^{(s)} \cdot [\![t_{n-\ell},\dots,t_n]\!]\partial_t^\rho\phi - \partial_t^{\rho+\ell}\phi(t_n) \right\|_B \le C\, |\phi|_{C^{\rho+m}([t_{n-\ell},t_n],B)}\, \Delta_n^{m-\ell} \qquad \forall 0 \le \ell \le m,$$

so that the combination with (2) yields

$$\begin{aligned}
\left\| \sum_{\ell=0}^{m-1} \delta w_{n,\ell}^{\text{far}} \right\|_D &\le C \sum_{\ell=0}^{m-1} |\phi|_{C^{\rho+m}([t_{n-\ell},t_n],B)}\, \Delta_n^{m-\ell} \int_{\gamma_{\text{far}}} |z|^{\mu-\rho-\ell-1}\, dz \qquad (55) \\
&\le C\, |\phi|_{C^{\rho+m}([t_{n-m+1},t_n],B)}\, \Delta^{m+\rho-\mu}.
\end{aligned}$$

To estimate $w^{(n)}_{\mathrm{far},m}$, we substitute γ by γ_{far} in the right-hand side of (43), multiply by $\mathbf{e}^{(s)}\cdot$ from the left, and observe that "$\left(K_\rho\left(\partial_t^{\Theta}\right)\left(\partial_t^\rho\phi\right)\right)^{(n)}$" in (43) then has to be substituted by "$w^{(n)}_{\mathrm{far},\rho,m}$". From Proposition 11 and the proof of Theorem 13 we then deduce (cf. (44))

$$\begin{aligned}\left\|w^{(n)}_{\mathrm{far},m}\right\|_D &\le \sqrt{s}\frac{\left(\beta_0\left\|\mathbf{A}\right\|\right)^2}{2\pi}\sum_{k=0}^{n}\Delta_k\,\mathrm{e}^{C\sigma_0(t_n-t_k)}\left\|[\![t_{k-m},\dots,t_k]\!]\partial_t^\rho\phi\right\|_B\int_{\gamma_{\mathrm{far}}}|z|^{\mu-\rho-m}\,dz\\ &\le C\,\mathrm{e}^{c\sigma T}\,\Delta^{m+\rho-\mu-1}\left\|\phi\right\|_{C^{m+\rho}([0,T],B)}.\end{aligned}$$

The last term in (54), $w_{\mathrm{far},m}(t_n)$, can be estimated by using (51):

$$\left\|u_{\rho+m}(z,t_n)\right\|_B \le \left\|\phi\right\|_{C^{\rho+m}([0,T],B)}\int_0^{t_n}\mathrm{e}^{z(t_n-\tau)}\,d\tau \le \frac{\mathrm{e}^{\sigma T}}{|z|}\left\|\phi\right\|_{C^{\rho+m}([0,T],B)}$$

and, in turn,

$$\left\|w_{\mathrm{far},m}(t_n)\right\|_D \le C\,\mathrm{e}^{\sigma T}\left\|\phi\right\|_{C^{\rho+m}([0,T],B)}\Delta^{m+\rho-\mu}.$$

The estimate of the farfield follows by choosing $m=q+1$.

Near Field

Estimate of $\partial_t^k u(z,\cdot)$ in the nearfield.

It holds

$$u_\rho(z,t)=\int_0^t\mathrm{e}^{z\tau}\,\partial_t^\rho\phi(t-\tau)\,d\tau.$$

By differentiating this relation k times for some $k\le p+1$ we get

$$\partial_t^k u_\rho(z,t)=\int_0^t\mathrm{e}^{z(t-\tau)}\,\partial_t^{\rho+k}\phi(\tau)\,d\tau+\mathrm{e}^{zt}\sum_{\ell=0}^{k-1}z^{k-1-\ell}\partial_t^{\rho+\ell}\phi(0).$$

Hence, we obtain from the assumption of the theorem

$$\begin{aligned}\left\|\partial_t^k u_\rho(z,t)\right\|_B &\le \mathrm{e}^{\sigma T}\left(|z|^{-1}\left|\phi\right|_{C^{\rho+k}([0,T],B)}+\begin{cases}0 & k\le m\\ \displaystyle\sum_{\ell=m}^{k-1}|z|^{k-1-\ell}\,\partial_t^{\rho+\ell}\phi(0) & m+1\le k\le p+1\end{cases}\right)\\ &\le \mathrm{e}^{\sigma T}\,|z|^{-1-\min\{0,m-k\}}\left\|\phi\right\|_{C^{\rho+k}([0,T],B)}.\end{aligned}\tag{56}$$

Solving the error recursion.

In order to estimate

$$\frac{1}{2\pi\,\mathrm{i}}\int_{\gamma_{\mathrm{near}}}\left(K_\rho(z)\left(u_\rho(z,t_n)-u_\rho^{(n)}(z)\right)\right.dz,\tag{57}$$

we analyze the error

$$e_n(z) := u_\rho(z, t_n) - u_\rho^{(n)}(z), \qquad z \in \gamma_{\text{near}}. \tag{58}$$

Following [16, proof of Theorem 3.3], we set

$$d_i^{(n)}(z) = u_\rho(z, t_{n-1} + c_i\Delta_n) - u_\rho(z, t_{n-1}) - \Delta_n \sum_{j=1}^{s} a_{ij} u'_\rho(z, t_{n-1} + c_j\Delta_n),$$

$$d^{(n)}(z) = u_\rho\left(z, t_n\right) - u_\rho(z, t_{n-1}) - \Delta_n \sum_{j=1}^{s} b_j u'_\rho(z, t_{n-1} + c_j\Delta_n) = d_s^{(n)}.$$

We set $\mathbf{D}^{(n)} = (d_i^{(n)})_{i=1}^s$ and

$$\boldsymbol{\delta}^{(k)} := \left(\delta_i^{(k)}\right)_{i=1}^s := \frac{1}{(k-1)!}\left(\mathbf{A}\mathbf{c}^{(k-1)\odot} - \frac{1}{k}\mathbf{c}^{k\odot}\right).$$

By inserting the exact solution into the Runge–Kutta scheme and performing Taylor expansion around t_n we obtain

$$\begin{aligned} \mathbf{D}^{(n)}(z) &= \textstyle\sum_{k=q+1}^{p} \Delta_n^k \partial_t^k u_\rho\left(z, t_n\right) \boldsymbol{\delta}^{(k)} + \Delta_n^p \mathbf{Q}^{(n)}(z), \\ d^{(n)}(z) &= \Delta_n^p \textstyle\int_{t_{n-1}}^{t_n} \kappa\left(\frac{t-t_{n-1}}{\Delta_n}\right) \partial_t^{p+1} u_\rho(z, t)\, dt, \end{aligned} \tag{59}$$

where

$$\mathbf{Q}^{(n)}(z) := \int_{t_{n-1}}^{t_n} \boldsymbol{\kappa}\left(\frac{t-t_{n-1}}{\Delta_n}\right) \partial_t^{p+1} u_\rho(z, t)\, dt$$

and $\boldsymbol{\kappa} = (\kappa_i)_{i=1}^s$, κ are bounded Peano kernels. Note that this implies

$$\begin{aligned} \left\|\mathbf{Q}^{(n)}(z)\, dz\right\|_B &\le C\Delta_n \left|u_\rho\left(z, \cdot\right)\right|_{C^{p+1}([t_{n-1}, t_n], B)} \overset{(56)}{\le} C\, \mathrm{e}^{\sigma T}\, \Delta_n\, |z|^{p-m}\, \|\phi\|_{C^{\rho+p+1}([0,T], B)}, \\ \|d_n(z)\|_D &\le C\Delta_n^{p+1} \left|u_\rho\left(z, \cdot\right)\right|_{C^{p+1}([t_{n-1}, t_n], B)} \le C\, \mathrm{e}^{\sigma T}\, \Delta_n^{p+1}\, |z|^{p-m}\, \|\phi\|_{C^{\rho+p+1}([0,T], B)}. \end{aligned}$$

Thus, the error satisfies the recursion

$$e_n(z) = R(\Delta_n z) e_{n-1}(z) - \Delta_n z \mathbf{b} \cdot (\mathbf{I} - \Delta_n z \mathbf{A})^{-1} \mathbf{D}^{(n)}(z) + d^{(n)}(z),$$

for the stability function R of the Runge–Kutta method (11). Solving the recursion and using that $e_0 = 0$ we obtain

$$e_n(z) = \sum_{j=1}^{n} \left(\prod_{\ell=j+1}^{n} R(\Delta_\ell z)\right) \left(\Delta_j z \mathbf{b} \cdot (\mathbf{I} - \Delta_j z \mathbf{A})^{-1} \mathbf{D}^{(j)}(z) + d^{(j)}(z)\right).$$

By Lemma 12 for Δ small enough we can estimate

$$|R(\Delta_n z)| \le \mathrm{e}^{C\Delta_n \sigma}, \qquad \forall z \in \gamma,\ n \ge 1, \tag{60}$$

so that

$$\|e_n(z)\|_B \leq C\,\mathrm{e}^{\sigma T}\sum_{j=1}^{n}\left(\left\|\Delta_j z\mathbf{b}\cdot(\mathbf{I}-\Delta_j z\mathbf{A})^{-1}\mathbf{D}^{(j)}(z)\right\|_B+\left\|d^{(j)}(z)\right\|_B\right). \tag{61}$$

The combination of the order condition (15) with (59) allows to bound the first norm in the right-hand side of (61) by

$$\begin{aligned}\left\|\Delta_j z\mathbf{b}\cdot(\mathbf{I}-\Delta_j z\mathbf{A})^{-1}\mathbf{D}^{(j)}(z)\right\|_B \leq \sum_{k=q+1}^{p}\Delta_j^k\left\|\partial_t^k u_\rho(z,t_j)\right\|_B\left\|\Delta_j z\mathbf{b}\cdot(\mathbf{I}-\Delta_j z\mathbf{A})^{-1}\boldsymbol{\delta}^{(k)}\right\| \\ +\Delta_j^p\left\|\Delta_j z\mathbf{b}\cdot(\mathbf{I}-\Delta_j z\mathbf{A})^{-1}\mathbf{Q}^{(j)}\right\|_B.\end{aligned} \tag{62}$$

For sufficiently small $0<C_{\mathrm{split}}=\mathcal{O}(1)$ in (50) we have $\|\Delta_j z\mathbf{A}\|<1$ for all $z\in\gamma_{\mathrm{near}}$ so that a Neumann series argument gives us

$$\left\|\Delta_j z\mathbf{b}\cdot(\mathbf{I}-\Delta_j z\mathbf{A})^{-1}\boldsymbol{\delta}^{(k)}\right\|\leq\frac{(C\Delta_j|z|)^{p-k+2}}{(k-1)!}$$

where C depends on $\mathbf{A},\mathbf{b},\mathbf{c}$. Recall that $m\leq q+1$. Thus, for all $z\in\gamma_{\mathrm{near}}$ it holds (cf. (56))

$$\begin{aligned}&\sum_{k=q+1}^{p}\Delta_j^k\left\|\partial_t^k u_\rho(z,t_j)\right\|_B\left\|\Delta_j z\mathbf{b}\cdot(\mathbf{I}-\Delta_j z\mathbf{A})^{-1}\boldsymbol{\delta}^{(k)}\right\|_B\\ &\qquad\leq C\sum_{k=q+1}^{p}\Delta_j^k\left\|\partial_t^k u_\rho(z,t_j)\right\|_B\frac{(C\Delta_j|z|)^{p-k+2}}{(k-1)!}\\ &\qquad\leq C_p\,\mathrm{e}^{\sigma T}\Delta_j^{p+2}|z|^{p+1-m}\|\phi\|_{C^{\rho+p}([0,T],B)}.\end{aligned}$$

For the second term in the right-hand side of (62) we get in a similar fashion

$$\Delta_j^p\left\|\Delta_j z\mathbf{b}\cdot(\mathbf{I}-\Delta_j z\mathbf{A})^{-1}\mathbf{Q}^{(j)}\right\|_B\leq C\Delta_j^p\left\|\mathbf{Q}^{(j)}\right\|_B\leq C\,\mathrm{e}^{\sigma T}\Delta_j^{p+1}|z|^{p-m}\|\phi\|_{C^{\rho+p+1}([0,T],B)},$$

so that

$$\|e_n(z)\|_B\leq C\,\mathrm{e}^{2\sigma T}\left(\Delta_j^{p+1}|z|^{p+1-m}\|\phi\|_{C^{\rho+p}([0,T],B)}+\Delta^p|z|^{p-m}\|\phi\|_{C^{\rho+p+1}([0,T],B)}\right).$$

This estimate allows to bound the nearfield error by using (2)

$$\begin{aligned}\|\delta\phi_n^{\mathrm{near}}\|_B&\leq C\int_{\gamma_{\mathrm{near}}}|z|^{\mu-\rho}\|e_n(z)\|_D\,dz\\ &\leq C\,\mathrm{e}^{2\sigma T}\|\phi\|_{C^{\rho+p+1}([0,T],D)}\int_{\gamma_{\mathrm{near}}}\left(\Delta_j^{p+1}|z|^{\mu-\rho+p+1-m}+\Delta^p|z|^{\mu-\rho+p-m}\right)dz\\ &\leq C\,\mathrm{e}^{2\sigma T}\|\phi\|_{C^{\rho+p+1}([0,T],D)}\begin{cases}\Delta^p & \mu-\rho+p-m<-1,\\ \Delta^p\log\frac{1}{\Delta} & \mu-\rho+p-m=-1,\\ \Delta^{m+\rho-1-\mu} & \mu-\rho+p-m>-1.\end{cases}\end{aligned}$$

The combination with the farfield estimates leads to the assertion for $\mu-\rho<-1$.

Case 2: $\mu-\rho\geq -1$.

Let $\nu\in\mathbb{N}_0$ be the smallest integer such that $\nu>\mu+1$ holds. Then the contour integral in

$$w=\frac{1}{2\pi\,\mathrm{i}}\int_{\gamma}(K_\nu(z)u_\nu(z,t)\,dz$$

is well defined for all $\phi\in C_0^\nu\left([0,T],D\right)$. Since ν is large enough we may choose γ as any suitable contour in the complex plane: either a vertical contour $\gamma_\perp$ running from $\sigma-\mathrm{i}\,\infty$ to $\sigma+\mathrm{i}\,\infty$ or a suitable closed contour $\gamma_{\circlearrowleft}$ clockwise oriented.

The representation of the discrete solution

$$w_\rho^{(n)}=\frac{1}{2\pi\,\mathrm{i}}\int_{\gamma_{\circlearrowleft}}\left(K^{-1}\right)_\rho(z)\,u_\rho^{(n)}(z)\;dz=\frac{1}{2\pi\,\mathrm{i}}\int_{\gamma_{\circlearrowleft}}\left(K^{-1}\right)_\nu(z)\,z^{\nu-\rho}u_\rho^{(n)}(z)\;dz$$

is well defined by Theorem 13, (42) if we choose a closed contour $\gamma_{\circlearrowleft}$ which encircles the spectra $\bigcup_{k=1}^{N}\sigma\left(\Delta_k^{-1}\mathbf{A}^{-1}\right)$. The error at time step t_n is given by

$$w\left(t_n\right)-w_\rho^{(n)}=\frac{1}{2\pi\,\mathrm{i}}\int_{\gamma_{\circlearrowleft}}(K^{-1})_\nu(z)\left(u_\nu(z,t_n)-z^{\nu-\rho}u_\rho^{(n)}(z)\right)\,dz. \tag{63}$$

By adding and subtracting $u_\nu^{(n)}$ we can split the error into two terms

$$\begin{aligned}T_1^{(n)}&=\frac{1}{2\pi\,\mathrm{i}}\int_{\gamma_\perp}\left(K^{-1}\right)_\nu(z)\left(u_\nu(z,t_n)-u_{n,\nu}(z)\right)\,dz,\\ T_2^{(n)}&=\frac{1}{2\pi\,\mathrm{i}}\int_{\gamma_{\circlearrowleft}}\left(K^{-1}\right)_\nu(z)\left(u_\nu^{(n)}(z)-z^{\nu-\rho}u_\rho^{(n)}(z)\right)\,dz.\end{aligned}$$

The term T_1 can be estimated by using Case 1 with the substitution $\rho\leftarrow\nu$ therein and we get

$$\left\|T_{n,1}\right\|_B\leq C\left\|\phi\right\|_{C^{\nu+p+1}([0,T],D)}c_{\mu-\nu+p-m}\left(\Delta\right)\Delta^{\min\{p,\nu+m-1-\mu\}}. \tag{64}$$

Note that $T_2^{(n)}$ is the s-th component of

$$\mathbf{T}_2^{(n)}=\left(\left(K^{-1}\right)_\nu\left(\partial_t^\Theta\right)\left(\partial_t^\nu\phi-[\![\Theta]\!]^{\nu-\rho}\partial_t^\rho\phi\right)\right)^{(n)}.$$

Theorem 13 for the choices $\tilde{m}\leftarrow 0$ and $\tilde{\rho}\leftarrow\nu$ can be applied since

$$\mu-\tilde{\rho}+1<\tilde{m}<q+1$$

so that

$$\left\|\mathbf{T}_2^{(n)}\right\|_B\leq C\sum_{k=0}^{n}\Delta_k\,\mathrm{e}^{C\sigma_0(t_n-t_k)}\left\|\partial_t^{\nu-\rho}\left(\partial_t^\rho\phi\right)^{(k)}-[\![t_{k-(\nu-\rho)},\ldots,t_k]\!]\partial_t^\rho\phi\right\|_D.$$

Proposition 11 leads to

$$\left\|\mathbf{T}_2^{(n)}\right\|_B \leq C\sum_{k=0}^{n}\Delta_k\,\mathrm{e}^{C\sigma_0(t_n-t_k)}\,|\partial_t^\rho\phi|_{C^{q+1}\left(\left[t_{k-(\nu-\rho)},t_k\right],D\right)}\,\Delta_k^{q+\rho+1-\nu} \tag{65}$$
$$\leq C\,\mathrm{e}^{\sigma T}\,|\partial_t^\rho\phi|_{C^{q+1}([0,T],D)}\,\Delta^{q+1-(\nu-\rho)}.$$

We choose $m=q+1$ in (64) and, since in this Case 2 we have

$$\rho\leq\mu+1<\nu,$$

the Δ-exponents in (64) and (65) satisfy

$$\nu+m-1-\mu=q+\nu-\mu>q+1\geq q+1-(\nu-\rho)\,.$$

This leads to the final error estimate

$$\left\|w\left(t_n\right)-w_\rho^{(n)}\right\|_D\leq C\,\mathrm{e}^{\sigma T}\,\|\phi\|_{C^{\nu+p+1}([0,T],B)}\,\Delta^{\min\{p,q+1+\rho-\nu\}}.$$

■

5 Runge-Kutta Generalized Convolution Quadrature for Solving Convolution Equations

5.1 Discretization

In this section we will consider the *solution* of one-sided convolution equations: For given g, find ϕ

$$K\left(\partial_t\right)\phi=g. \tag{66}$$

We assume that the transfer operator K satisfies

$$K\in\mathcal{A}_{\sigma_+}^{\theta}\left(B,D\right)\qquad\text{for some }\sigma_+,\theta\in\mathbb{R} \tag{67a}$$

and, in analogy to (4), we choose $m\in\mathbb{N}_0$ as the smallest integer such that $m>\theta+1$. In view of (6) we are seeking the solution ϕ of (66) in $C_0^m\left([0,T]\,,B\right)$.

To ensure existence of a solution of (66) we assume

$$K^{-1}:\mathbb{C}_{\sigma_-}\to\mathcal{L}\left(D,B\right)\ \text{exists and}\ K^{-1}\in\mathcal{A}_{\sigma_-}^{\mu}\left(D,B\right)\ \text{for some}\ \sigma_-,\mu\in\mathbb{R}. \tag{67b}$$

We define ν according to (4) but emphasize that μ, this time, denotes the growth exponent of the *inverse* operator K^{-1}.

Proposition 15 *Let (67) be satisfied. If $g\in C_0^\nu\left([0,T]\,,D\right)$, then*

$$\phi\left(t\right):=\left(K^{-1}\left(\partial_t\right)g\right)\left(t\right)=\frac{1}{2\pi\,\mathrm{i}}\int_\gamma\left(K^{-1}\right)_\nu\left(z\right)\left(\int_0^t\mathrm{e}^{z\tau}\,\partial_t^\nu g\left(t-\tau\right)d\tau\right)dz \tag{68}$$

for a contour $\gamma=\sigma+\mathrm{i}\,\mathbb{R}$ and $\sigma>\sigma_-$ is well defined.

If $g\in C_0^{\nu+m}\left([0,T]\,,D\right)$, it holds $\phi\in C_0^m\left([0,T]\,,B\right)$ so that $K\left(\partial_t\right)\phi$ is well defined and ϕ as in (68) satisfies (66).

Proof. The choice of ν and the smoothness assumption on g imply that ϕ in (68) is well defined (cf. (7)). By differentiating (68) and using $g \in C_0^{\nu+m}([0,T],D)$, we obtain $\phi^{(r)} = 0$ for $0 \le r \le m-1$. Thus, the associativity for one-sided convolutions (see [14, (2.3), (2.22)])

$$V(\partial_t) W(\partial_t) = (VW)(\partial_t) \tag{69}$$

yields $K(\partial_t)\left(K^{-1}(\partial_t)\phi\right) = g$. ■

The inversion formula (68) allows us to discretize the convolution equation (66) by the same method as developed for the forward equation (cf. Section 3):

$$\bigtimes_{n=1}^{N} \boldsymbol{\phi}_\rho^{(n)} := \left(K^{-1}\right)_\rho \left(\partial_t^\Theta\right) \partial_t^\rho g \quad \text{for some } \rho \text{ as in (16)} \tag{70a}$$

and the approximation of ϕ at time point t_n is given by the last component

$$\phi(t_n) \approx \phi_\rho^{(n)} := \mathbf{e}^{(s)} \cdot \boldsymbol{\phi}_\rho^{(n)}. \tag{70b}$$

Remark 16 *The representation of the generalized convolution quadrature in the form (70) is well suited for theoretical investigations but not for the practical implementation: For important applications such as, e.g., for the solution of the space-time wave equation, the operator $K^{-1}(s)$ is infinite dimensional and not available explicitly so that its discretization would be prohibitive expensive. Instead, we will prove that the associativity of continuous convolutions (69) is inherited by the* Runge-Kutta gCQ*: Under assumptions which will be detailed in Theorem 26 it holds*

$$V\left(\partial_t^\Theta\right) \circ W\left(\partial_t^\Theta\right) = (VW)\left(\partial_t^\Theta\right) \tag{71}$$

so that (70a) can be written in the form (cf. Remark 20, Corollary 27)

$$K_{-\rho}\left(\partial_t^\Theta\right)\left(\bigtimes_{n=1}^{N} \boldsymbol{\phi}_\rho^{(n)}\right) = \bigtimes_{n=1}^{N} \left(\partial_t^\rho \mathbf{g}\right)^{(n)}.$$

Definition 17 (Runge-Kutta gCQ for Solving Convolution Equations) *Let the transfer operator K satisfy (67) and let $\nu, m \in \mathbb{N}_0$ be the smallest integers such that $\nu > \mu + 1$ and $m > \theta + 1$. Let $g \in C_0^{\nu+m}([0,T],D)$. We consider the problem: Find $\phi \in C_0^m([0,T],B)$ such that*

$$K(\partial_t)\phi = g. \tag{72}$$

Let a Runge-Kutta method be given which satisfies Assumption 4. Then the discretization of (72) by Runge-Kutta generalized Convolution Quadrature *is given by*

$$K_{-\rho}\left(\partial_t^\Theta\right)\left(\bigtimes_{n=1}^{N} \boldsymbol{\phi}_\rho^{(n)}\right) = \bigtimes_{n=1}^{N} \left(\partial_t^\rho \mathbf{g}\right)^{(n)} \tag{73}$$

and the approximation of ϕ at time t_n by the last component $\phi_\rho^{(n)} := \mathbf{e}^{(s)} \cdot \boldsymbol{\phi}_\rho^{(n)}$. Here, $\rho \in \mathbb{N}_0$ is a regularization parameter which can be chosen in the range

$$\nu - (q+1) \leq \rho \leq p + \nu - (q+1)\,, \tag{74}$$

where p denotes de order and q the stage order of the Runge-Kutta method.

Remark 18 *For the algorithmic realization of the Runge-Kutta gCQ (cf. (73)) one has to approximate the contour integrals in*

$$\frac{1}{2\pi\,\mathrm{i}} \int_\gamma z^\rho K(z)\, \mathbf{u}_\rho^{(n)}(z)\, dz \tag{75}$$

by numerical quadrature. For the implicit Euler gCQ, such a quadrature scheme has been proposed and analyzed in [9, 10].

On one hand, Theorem 14 indicates that the upper bound in (74) for the choice of ρ improves the convergence rates up to the optimal order $\mathcal{O}(\Delta^p)$ for sufficiently smooth and compatible data, while smaller choices of ρ lead to a milder growth behavior of the integrand in (75) and simplify the numerical quadrature. This also shows the importance of the summation-by-parts representation which allows to achieve a faster decay of the integrand in the error estimates without increasing the numerical parameter ρ furthermore.

5.2 Associativity

The stability and convergence analysis of the approximation $\phi_\rho^{(n)}$ as in Definition 17 follows directly from Theorem 13 and 14 if we prove the inversion formula

$$\bigtimes_{n=1}^{N} \boldsymbol{\phi}_\rho^{(n)} = \left(K^{-1}\right)_\rho \left(\partial_t^\Theta\right) \left(\bigtimes_{n=1}^{N} \left(\partial_t^\rho \mathbf{g}\right)^{(n)} \right).$$

In more generality, we will prove (71). This requires to reformulate the contour integrals via *tensorial divided differences* which we will introduce and the proof of a Leibniz rule for tensorial divided differences to derive the associativity property for the composition of discrete generalized convolution operators. We refer to [7] and [6] for an introduction to tensor calculus and advanced topics.

For $i, j, i', j' \in \{1, \ldots, N\}$, we consider sequences $\mathbf{B}^{(k)} \in \mathbb{C}^{s\times s}$, $k \in \{i, \ldots, j\}$ and $\mathbf{C}^{(k)} \in \mathbb{C}^{s\times s}$, $k \in \{i', \ldots, j'\}$, of matrices. In Section 4.1 we introduced the Kronecker products of matrices and their application to tensors of vectors. The *composition* of Kronecker matrices is defined as the tensor of the "matching" matrix products by

$$\left(\bigotimes_{k=i}^{j} \mathbf{B}^{(k)} \right) \circ \left(\bigotimes_{k=i'}^{j'} \mathbf{C}^{(k)} \right) = \prod_{k=\min\{i,i'\}}^{\max\{j,j'\}} \mathbf{B}^{(k)} \mathbf{C}^{(k)},$$

where we set $\mathbf{B}^{(k)} = \mathbf{I}$ for $k \notin \{i, \ldots, j\}$ and $\mathbf{C}^{(k)} = \mathbf{I}$ for $k \notin \{i', \ldots, j'\}$. For $i = i'$ and $j = j'$ we suppress the composition sign "$\circ$" as is usual for matrix-matrix multiplication.

Finally we define the resolvent matrix for $\mathbf{C} \in \mathbb{C}^{s\times s}$ by

$$\mathbf{R}_z(\mathbf{C}) \in \mathbb{C}^{s\times s} \quad \text{with} \quad \mathbf{R}_z(\mathbf{C}) := (z\mathbf{I} - \mathbf{C})^{-1}.$$

Definition 19 *For a set of matrices $\mathbf{C}^{(k)} \in \mathbb{C}^{s\times s}$, $1 \le k \le n$, and a function f which is analytic in a complex neighborhood $\mathcal{U}$ of $\bigcup_{k=1}^{n} \sigma\left(\mathbf{C}^{(k)}\right)$, the* tensorial divided difference $\left[\times_{k=1}^{n} \mathbf{C}^{(k)}\right] f$ *is a Kronecker matrix given by*[2]

$$\left[\bigtimes_{k=1}^{n} \mathbf{C}^{(k)}\right] f := \frac{1}{2\pi\,\mathrm{i}} \int_\Gamma f(z) \left(\bigotimes_{k=1}^{n} \mathbf{R}_z\left(\mathbf{C}^{(k)}\right)\right) dz, \tag{76}$$

for a counterclockwise oriented closed contour Γ in $\mathcal{U}$ which encircles $\bigcup_{k=1}^{n} \sigma\left(\mathbf{C}^{(k)}\right)$.

Tensorial divided differences $\left[\times_{k=i}^{j} \mathbf{C}^{(k)}\right] f$ are generalizations of standard divided differences for 1×1 matrices $\mathbf{C}^{(k)} = (x_k)$ with nodal points x_k: In the latter case, divided differences allow for a contour integral representation (cf. Remark 21) which is generalized by (76) for the case of matrices $\mathbf{C}^{(k)}$. In Section 23 we will derive an alternative representation of tensorial divided differences which mimics the recurrence relation for classical divided differences.

These tensorial divided differences allow to express the generalized discrete convolution (22), (25) via

$$\begin{aligned}\boldsymbol{\phi}_\rho^{(n)} &= \left(\left(K^{-1}\right)_\rho \left(\partial_t^\Theta\right) \partial_t^\rho g\right)^{(n)} \\ &= \sum_{k=1}^{n} \omega_{n,k}(0) \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \left(\left[\bigtimes_{\ell=k}^{n} \frac{\mathbf{A}^{-1}}{\Delta_\ell}\right] \left(K^{-1}\right)_\rho \left(\partial_t^\rho \mathbf{g}^{(k)} \otimes \left(\mathbf{A}^{-1}\mathbf{1}\right)^{(n-k)\otimes}\right)\right),\end{aligned} \tag{77}$$

for $1 \le n \le N$. The result is an N-tuple of vectors in $\mathbb{C}^s$.

The function $\omega_{n,j}$ is given by

$$\omega_{n,j}(z) := \prod_{\ell=j+1}^{n} \left(z - \Delta_\ell^{-1}\right). \tag{78}$$

Remark 20 *This representation shows that the generalized discrete convolution depends only on the discrete values $\partial_t^\rho \mathbf{g}^{(k)}$ and thus can be applied also to tuples $\times_{\ell=1}^{N} \left(\partial_t^\rho \mathbf{g}\right)^{(k)}$ of stage vectors; thus, the composition of generalized discrete convolutions is well defined.*

[2]We prefer the notation $\left[\times_{k=1}^{n} \mathbf{C}^{(k)}\right] f$ instead of $\left[\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \ldots, \mathbf{C}^{(n)}\right] f$ because of brevity.

The representation (78) extends the definition of generalized discrete convolutions for the implicit Euler method (cf. [10]) to Runge-Kutta methods as can be seen from the following remark.

Remark 21 *In [11, First formula in the proof of Lemma 4.1.], it was shown that the gCQ based on the implicit Euler method with variable step size can be written in the form*

$$\phi_\rho^{(n)} = \sum_{j=1}^{n} \omega_{n,j}\left(0\right) \left(\left[\frac{1}{\Delta_j}, \frac{1}{\Delta_{j+1}}, \ldots, \frac{1}{\Delta_n} \right] \left(K^{-1}\right)_\rho \right) \partial_t^\rho g^{(j)}, \tag{79}$$

where $\omega_{n,k}$ is as in (78). Note that the divided differences of an analytic function f have the following contour integral representation

$$\left[x_1, x_2, \ldots, x_N\right] f = \frac{1}{2\pi\,\mathrm{i}} \int_C \frac{f\left(z\right)}{\prod_{i=1}^{N} \left(z - x_i\right)} dz,$$

for a counterclockwise oriented contour C enclosing the arguments x_i, $i = 1, \ldots, N$. Hence, taking into account the clockwise orientation of the contour γ, (79) can be expressed in terms of contour integrals as

$$\phi_\rho^{(n)} = \sum_{j=1}^{n} \Delta_j \frac{1}{2\pi\,\mathrm{i}} \int_\gamma \left(\prod_{\ell=j}^{n} \frac{1}{1 - z\Delta_\ell} \right) \left(K^{-1}\right)_\rho \left(z\right) \partial_t^\rho g^{(j)} dz. \tag{80}$$

Alternatively, we consider equation (77) for the implicit Euler method. In this case we have $\mathbf{A} = (1) \in \mathbb{R}^{1\times 1}$ and, in turn,

$$\begin{aligned} \phi_\rho^{(n)} &\overset{(77)}{=} \sum_{k=1}^{n} \omega_{n,k}\left(0\right) \left[\bigtimes_{\ell=k}^{n} \Delta_\ell^{-1} \right] \left(K^{-1}\right)_\rho \partial_t^\rho g^{(k)} \\ &\overset{(76)}{=} \sum_{k=1}^{n} \Delta_k \frac{1}{2\pi\,\mathrm{i}} \int_\gamma \left(\prod_{\ell=k}^{n} \frac{1}{1 - z\Delta_\ell} \right) \left(K^{-1}\right)_\rho \left(z\right) \partial_t^\rho g^{(k)} dz. \end{aligned}$$

This is the same expression as (80) and we see that (77) defines an extension of the divided difference representation of scalar *generalized convolution operators for the implicit Euler method to Runge-Kutta methods.*

The key role for writing (70a) as a forward equation will be played by an elegant inversion formula (which is well known for Runge-Kutta Convolution Quadrature with *constant* time steps).

In order to prove the associativity property of our discretization we develop a *tensorial* Leibniz formula and a composition rule for tensorial divided differences.

By Cauchy's integral theorem it is easy to see that $[\mathbf{C}]\, f$ is the value of the function f applied to the matrix $\mathbf{C}$ which is the analogue to standard zero-th

order divided differences. For higher order divided differences we first introduce the tensorial difference $\ominus^{(k,j)}(\mathbf{A},\mathbf{B})$ as the Kronecker matrix defined by

$$\ominus^{(k,j)}(\mathbf{A},\mathbf{B}) = \left(\bigotimes_{\ell=1}^{k-1}\mathbf{I}\right)\otimes\mathbf{A}\otimes\bigotimes_{\ell=k+1}^{n}\mathbf{I} - \left(\bigotimes_{\ell=1}^{j-1}\mathbf{I}\right)\otimes\mathbf{B}\otimes\bigotimes_{\ell=j+1}^{n}\mathbf{I},$$

If $\mathbf{A}$ and $\mathbf{B}$ are simultaneously diagonalizable, this is, $\mathbf{A} = \mathbf{V}^{-1}\mathbf{D}^{(1)}\mathbf{V}$ and $\mathbf{B} = \mathbf{V}^{-1}\mathbf{D}^{(2)}\mathbf{V}$, for some $\mathbf{V}$ and diagonal matrices $\mathbf{D}^{(1)}$, $\mathbf{D}^{(2)}$, we have[3]

$$\left(\bigotimes_{i=1}^{n}\mathbf{v}^{(i)}\right)\cdot\left(\ominus^{(k,j)}(\mathbf{A},\mathbf{B})\bigotimes_{i=1}^{n}\mathbf{w}^{(i)}\right) = \left(\bigotimes_{i=1}^{n}\mathbf{V}^{-\intercal}\mathbf{v}^{(i)}\right)\cdot\left(\ominus^{(k,j)}\left(\mathbf{D}^{(1)},\mathbf{D}^{(2)}\right)\bigotimes_{i=1}^{n}\mathbf{V}\mathbf{w}^{(i)}\right).$$

Remark 22 *The eigenvalues of* $\ominus^{(k,j)}(\mathbf{A},\mathbf{B})$ *are given by* $\lambda_{i_1}^{(1)}-\lambda_{i_2}^{(2)}$*, where* $\lambda_{i_1}^{(1)}$ *are the eigenvalues of* $\mathbf{A}$ *and* $\lambda_{i_2}^{(2)}$ *those of* $\mathbf{B}$*. Hence,* $\ominus^{(k,j)}(\mathbf{A},\mathbf{B})$ *is regular if and only if* $\sigma(\mathbf{A})\cap\sigma(\mathbf{B}) = \emptyset$*. In this case,* $\left(\ominus^{(k,j)}(\mathbf{A},\mathbf{B})\right)^{-1}$ *exists, i.e.,*

$$\left(\ominus^{(k,j)}(\mathbf{A},\mathbf{B})\right)^{-1}\ominus^{(k,j)}(\mathbf{A},\mathbf{B}) = \ominus^{(k,j)}(\mathbf{A},\mathbf{B})\left(\ominus^{(k,j)}(\mathbf{A},\mathbf{B})\right)^{-1} = \bigotimes_{i=1}^{n}\mathbf{I}$$

but, in general, is not a Kronecker matrix. Further note that

$$\begin{aligned}\left(\bigotimes_{i=1}^{n}\mathbf{v}^{(i)}\right)\cdot\left(\left(\ominus^{(k,j)}(\mathbf{A},\mathbf{B})\right)^{-1}\bigotimes_{i=1}^{n}\mathbf{w}^{(i)}\right)\\ = \left(\bigotimes_{i=1}^{n}\mathbf{V}^{-\intercal}\mathbf{v}^{(i)}\right)\cdot\left(\left(\ominus^{(k,j)}\left(\mathbf{D}^{(1)},\mathbf{D}^{(2)}\right)\right)^{-1}\bigotimes_{i=1}^{n}\mathbf{V}\mathbf{w}^{(i)}\right).\end{aligned}$$

Lemma 23 *For a set of matrices* $\mathbf{C}^{(k)}\in\mathbb{C}^{s\times s}$*,* $1\le k\le n$*, which are simultaneously diagonalizable, i.e.,*

$$\mathbf{C}^{(k)} = \mathbf{V}^{-1}\mathbf{D}^{(k)}\mathbf{V}, \tag{81}$$

it holds

$$\left[\bigtimes_{k=1}^{n}\mathbf{C}^{(k)}\right]f = \left(\bigotimes_{k=1}^{n}\mathbf{V}^{-1}\right)\left(\left[\bigtimes_{k=1}^{n}\mathbf{D}^{(k)}\right]f\right)\left(\bigotimes_{k=1}^{n}\mathbf{V}\right). \tag{82}$$

Furthermore, if the intersection of the spectra of any pair $\mathbf{C}^{(k)},\mathbf{C}^{(j)}$*,* $k\neq j$*, is empty, the following recursion for tensorial divided differences holds true*

$$\begin{aligned}&\left[\mathbf{C}^{(1)},\ldots,\mathbf{C}^{(k)}\right]f \\ &\quad = \left(\left(\mathbf{I}\otimes\left[\mathbf{C}^{(2)},\ldots,\mathbf{C}^{(k)}\right]f\right) - \left(\left[\mathbf{C}^{(1)},\ldots,\mathbf{C}^{(k-1)}\right]f\otimes\mathbf{I}\right)\right)\left(\ominus^{(k,1)}\left(\mathbf{C}^{(k)},\mathbf{C}^{(1)}\right)\right)^{-1}.\end{aligned} \tag{83}$$

[3] By $\mathbf{V}^{\intercal}$ we denote the transposed of the matrix $\mathbf{V}$ (without complex conjugation) and by $\mathbf{V}^{-\intercal} = (\mathbf{V}^{-1})^{\intercal}$.

Proof. Statement (82) is trivial.

Since the matrices $\mathbf{C}^{(k)}$ are simultaneously diagonalizable it is sufficient to prove the statement for diagonal matrices $\mathbf{C}^{(k)} = \mathbf{D}^{(k)}$ and the statement follows from the corresponding property for standard divided differences. ■

Lemma 24 (Leibniz Rule for Tensorial Divided Differences) *Let* $\mathbf{C}^{(j)}$, $1 \leq j \leq n$, *and* f *be as in Definition 19. For mappings* f, g *analytic in a neighborhood of* $\bigcup_{k=1}^{n} \sigma(\mathbf{C}^{(k)})$ *the tensorial Leibniz' rule for divided differences holds*

$$\left[\bigtimes_{k=1}^{n} \mathbf{C}^{(k)}\right](fg) = \sum_{j=1}^{n} \left(\left[\bigtimes_{k=j}^{n} \mathbf{C}^{(k)}\right] f\right) \circ \left(\left[\bigtimes_{k=1}^{j} \mathbf{C}^{(k)}\right] g\right). \tag{84}$$

Proof. Since the matrices $\mathbf{C}^{(k)}$ are assumed to be simultaneously diagonalizable it is sufficient to prove the statement for diagonal matrices $\mathbf{C}^{(k)} = \mathbf{D}^{(k)}$, $1 \leq k \leq n$. Furthermore, continuity of divided differences with respect to the arguments $\mathbf{C}^{(k)}$, $1 \leq k \leq n$, implies that it is enough to prove (84) for matrices with pairwise disjoint spectra, cf. [4].

The statement is trivial for $n = 1$ and we assume next that the assertion holds for all $m < n$ and derive it for n.

From Lemma 23, we deduce[4]

$$\begin{aligned}
&\left[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(n)}\right](fg) \\
&= \left(\left(\mathbf{I} \otimes \left[\mathbf{D}^{(2)}, \ldots, \mathbf{D}^{(n)}\right](fg)\right) - \left(\left[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(n-1)}\right](fg) \otimes \mathbf{I}\right)\right) \left(\ominus^{(n,1)}\left(\mathbf{D}^{(n)}, \mathbf{D}^{(1)}\right)\right)^{-1} \\
&\overset{\text{ind. assump.}}{=} \left(\mathbf{I} \otimes \sum_{j=2}^{n} \left(\left[\mathbf{D}^{(j)}, \ldots, \mathbf{D}^{(n)}\right] f\right) \circ \left(\left[\mathbf{D}^{(2)}, \ldots, \mathbf{D}^{(j)}\right] g\right)\right. \\
&\quad \left. - \left(\sum_{j=1}^{n-1} \left[\mathbf{D}^{(j)}, \ldots, \mathbf{D}^{(n-1)}\right] f \circ \left[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(j)}\right] g\right) \otimes \mathbf{I}\right) \left(\ominus^{(n,1)}\left(\mathbf{D}^{(n)}, \mathbf{D}^{(1)}\right)\right)^{-1} \\
&= \left(\sum_{j=2}^{n} \left(\left[\mathbf{D}^{(j)}, \ldots, \mathbf{D}^{(n)}\right] f \circ \left[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(j)}\right] g\right) \ominus^{(j,1)}\left(\mathbf{D}^{(j)}, \mathbf{D}^{(1)}\right)\right. \\
&\quad \left. + \sum_{j=1}^{n-1} \left(\left[\mathbf{D}^{(j)}, \ldots, \mathbf{D}^{(n)}\right] f \circ \left[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(j)}\right] g\right) \ominus^{(n,j)}\left(\mathbf{D}^{(n)}, \mathbf{D}^{(j)}\right)\right) \frac{1}{\ominus^{(n,1)}\left(\mathbf{D}^{(n)}, \mathbf{D}^{(1)}\right)}.
\end{aligned}$$

[4] To derive the third equality, we have inserted

$$0 = -\sum_{j=2}^{n} \left[\mathbf{D}^{(j)}, \ldots, \mathbf{D}^{(n)}\right] f \circ \left(\left[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(j-1)}\right] g \otimes \mathbf{I}\right) + \sum_{j=1}^{n-1} \left(\mathbf{I} \otimes \left[\mathbf{D}^{(j+1)}, \ldots, \mathbf{D}^{(n)}\right] f\right) \circ \left[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(j)}\right] g$$

and used (83).

Since $\ominus^{(1,1)}\left(\mathbf{D}^{(1)},\mathbf{D}^{(1)}\right) = \ominus^{(n,n)}\left(\mathbf{D}^{(n)},\mathbf{D}^{(n)}\right) = 0$ the first sum can be extended to $j=1$ and the second one to $j=n$ without changing the values. Since $\ominus^{(j,1)}\left(\mathbf{D}^{(j)},\mathbf{D}^{(1)}\right)+\ominus^{(n,j)}\left(\mathbf{D}^{(n)},\mathbf{D}^{(j)}\right) = \ominus^{(n,1)}\left(\mathbf{D}^{(n)},\mathbf{D}^{(1)}\right)$, the result follows. ■

Finally, we will need a result for the composition of tensorized bilinear forms.

Lemma 25 *For vectors $\mathbf{v}^{(j)},\mathbf{w}^{(j)} \in \mathbb{C}^s$, let*

$$\mathbf{q}^{(k+1)} := \alpha^{(m+1,k)}\mathbf{B}^{(k+1)}\mathbf{w}^{(k+1)} \quad \textit{with} \quad \alpha^{(m+1,k)} := \left(\bigotimes_{j=m+1}^{k}\mathbf{v}^{(j)}\right)\cdot\left(\bigotimes_{j=m+1}^{k}\mathbf{B}^{(j)}\right)\bigotimes_{j=m+1}^{k}\mathbf{w}^{(j)}.$$

Then

$$\begin{aligned}&\left(\bigotimes_{\ell=k+1}^{n}\mathbf{v}^{(\ell)}\otimes\bullet\right)\cdot\left(\bigotimes_{\ell=k+1}^{n+1}\mathbf{C}^{(\ell)}\right)\left(\mathbf{q}^{(k+1)}\otimes\bigotimes_{\ell=k+2}^{n+1}\mathbf{w}^{(j)}\right) \\ &\qquad = \left(\bigotimes_{\ell=m+1}^{n}\mathbf{v}^{(\ell)}\otimes\bullet\right)\cdot\left(\bigotimes_{\ell=k+1}^{n+1}\mathbf{C}^{(\ell)}\right)\circ\left(\bigotimes_{\ell=m+1}^{k+1}\mathbf{B}^{(\ell)}\right)\bigotimes_{\ell=k+1}^{n+1}\mathbf{w}^{(j)}\end{aligned} \tag{85}$$

Proof. We denote the left-hand side in (85) by lhs. Then,

$$\begin{aligned}\text{lhs} &= \alpha^{(m+1,k)}\left(\bigotimes_{\ell=k+1}^{n}\mathbf{v}^{(\ell)}\otimes\bullet\right)\cdot\left(\bigotimes_{\ell=k+1}^{n+1}\mathbf{C}^{(\ell)}\right)\left(\mathbf{B}^{(k+1)}\mathbf{w}^{(k+1)}\otimes\bigotimes_{\ell=k+2}^{n+1}\mathbf{w}^{(j)}\right) \\ &= \alpha^{(m+1,k)}\left(\mathbf{v}^{(k+1)}\cdot\mathbf{C}^{(k+1)}\mathbf{B}^{(k+1)}\mathbf{w}^{(k+1)}\right)\left(\bigotimes_{\ell=k+2}^{n}\mathbf{v}^{(\ell)}\otimes\bullet\right)\cdot\left(\bigotimes_{\ell=k+2}^{n+1}\mathbf{C}^{(\ell)}\right)\bigotimes_{\ell=k+2}^{n+1}\mathbf{w}^{(j)} \\ &= \left(\left(\bigotimes_{j=m+1}^{k}\mathbf{v}^{(j)}\right)\cdot\left(\bigotimes_{j=m+1}^{k}\mathbf{B}^{(j)}\right)\bigotimes_{j=m+1}^{k}\mathbf{w}^{(j)}\right)\times \\ &\times\left(\mathbf{v}^{(k+1)}\cdot\mathbf{C}^{(k+1)}\mathbf{B}^{(k+1)}\mathbf{w}^{(k+1)}\right)\times \\ &\times\left(\bigotimes_{\ell=k+2}^{n}\mathbf{v}^{(\ell)}\otimes\bullet\right)\cdot\left(\bigotimes_{\ell=k+2}^{n+1}\mathbf{C}^{(\ell)}\right)\bigotimes_{\ell=k+2}^{n+1}\mathbf{w}^{(j)}\end{aligned}$$

and this is the assertion. ■

Theorem 26 (Associativity) *Let a Runge-Kutta method be given by the Butcher table $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ with non-singular $\mathbf{A}$. Let $W(s)\in\mathcal{L}(B,D)$ and $V(s)\in\mathcal{L}(D,E)$ denote transfer operators which are analytic in some complex neighborhood $\mathcal{U}$ of $\bigcup_{k=1}^{N}\sigma\left(\mathbf{M}^{(k)}\right)$. It holds*

$$V\left(\partial_t^{\Theta}\right)\circ W\left(\partial_t^{\Theta}\right) = (VW)\left(\partial_t^{\Theta}\right). \tag{86}$$

Proof. We set

$$\mathbf{q}^{(k+1)} := \left(\mathbf{e}^{(k-m)\otimes} \otimes \bullet\right) \left(\left[\bigtimes_{\ell=m}^{k} \frac{\mathbf{A}^{-1}}{\Delta_\ell}\right] W\right) \left(\mathbf{w}^{(m)} \times \left(\mathbf{A}^{-1}\mathbf{1}\right)^{(k-m)\times}\right).$$

The left-hand side in (86) can be written in the form

$$\begin{aligned}
&\left(V\left(\partial_t^\Theta\right)\left(W\left(\partial_t^\Theta\right)w\right)\right)^{(n)} \\
&\quad = \sum_{k=0}^{n}\sum_{m=0}^{k} \omega_{n,k}(0)\,\omega_{k,m}(0) \left(\mathbf{e}^{(n-k)\otimes} \otimes \bullet\right) \cdot \left[\bigtimes_{\ell=k}^{n} \frac{\mathbf{A}^{-1}}{\Delta_\ell}\right] V\left(\mathbf{q}^{(k+1)} \otimes \left(\mathbf{A}^{-1}\mathbf{1}\right)^{(n-k)\otimes}\right) \\
&\quad \overset{\text{Lem. 25}}{=} \sum_{m=0}^{n} \omega_{n,m}(0) \sum_{k=m}^{n} \left(\mathbf{e}^{(n-m)\otimes} \otimes \bullet\right) \cdot \\
&\qquad \cdot \left(\left[\bigtimes_{\ell=k}^{n} \frac{\mathbf{A}^{-1}}{\Delta_\ell}\right] V\right) \circ \left(\left[\bigtimes_{\ell=m}^{k} \frac{\mathbf{A}^{-1}}{\Delta_\ell}\right] W\right) \left(\mathbf{w}^{(m)} \otimes \left(\mathbf{A}^{-1}\mathbf{1}\right)^{(n-m)\otimes}\right).
\end{aligned}$$

Next we apply the tensorial Leibniz rule for divided differences (cf. Lemma 24) to obtain

$$\begin{aligned}
&\left(V\left(\partial_t^\Theta\right)\left(W\left(\partial_t^\Theta\right)w\right)\right)^{(n)} \\
&\quad = \sum_{m=0}^{n} \omega_{n,m}(0) \left(\mathbf{e}^{(n-m)\otimes} \otimes \bullet\right) \cdot \left(\left[\bigtimes_{\ell=m}^{n} \frac{\mathbf{A}^{-1}}{\Delta_\ell}\right] (VW)\right) \left(\mathbf{w}^{(m)} \otimes \left(\mathbf{A}^{-1}\mathbf{1}\right)^{(n-m)\otimes}\right) \\
&\quad = \left((VW)\left(\partial_t^\Theta\right)w\right)^{(n)}.
\end{aligned}$$

■

Corollary 27 (Inversion Formula) *Let a Runge-Kutta method be given by the Butcher table* $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ *with non-singular* $\mathbf{A}$. *Equation (70a) has an explicit inversion formula. It holds*

$$K_{-\rho}\left(\partial_t^\Theta\right)\left(\bigtimes_{n=1}^{N} \boldsymbol{\phi}^{(n)}\right) = \bigtimes_{n=1}^{N} \partial_t^\rho \mathbf{g}^{(n)}. \tag{87}$$

Proof. We employ Theorem 26 with $V := K_{-\rho}$ and $W := \left(K^{-1}\right)_\rho$ to obtain

$$\begin{aligned}
&\left(K_{-\rho}\left(\partial_t^\Theta\right)\left(\left(K^{-1}\right)_\rho\left(\partial_t^\Theta\right)w\right)\right)^{(n)} \\
&\quad = \sum_{m=0}^{n} \omega_{n,m}(0) \left(\mathbf{e}^{(n-m)\otimes} \otimes \bullet\right) \cdot \left[\bigtimes_{\ell=m}^{n} \frac{\mathbf{A}^{-1}}{\Delta_\ell}\right] (\mathrm{Id}) \left(\mathbf{w}^{(m)} \otimes \left(\mathbf{A}^{-1}\mathbf{1}\right)^{(n-m)\otimes}\right)
\end{aligned}$$

with the identity mapping Id. Hence, only the summand with $m = n$ is different from zero and the assertion follows. ■

6 Implementation and experiment

Our implementation of the Runge–Kutta gCQ is based on quadrature applied to definition (22). If a suitable quadrature with nodes z_ℓ and weights w_ℓ, $\ell = 1, \ldots, N_Q$, is available it is clear how to approximate *action* of the (forward) convolution $K(\partial_t^\Theta)\phi$ by a Runge-Kutta time stepping method applied to

$$\partial_t u_\rho(z,t) = z_\ell u_\rho(z,t) + \partial_t^\rho \phi; \quad u_\rho(z_\ell, 0) = 0, \quad \ell = 1, \ldots, N_Q.$$

The *solution* of the convolution equation $K(\partial_t^\Theta)\phi = g$, for given g, avoids the evaluation of the inverse convolution $\phi = K^{-1}(\partial_t^\Theta)g$ by employing the following algorithm which is based on K and not on its inverse. We compute approximations $\widetilde{\boldsymbol{\phi}}^{(n)} \approx \boldsymbol{\phi}^{(n)}$ from

$$K_{-\rho}\left((\Delta_n \mathbf{A})^{-1}\right)\widetilde{\boldsymbol{\phi}}^{(n)} = \mathbf{g}^{(n)} - \sum_{\ell=1}^{N_Q} w_\ell K_{-\rho}(z_\ell)\left(\mathbf{e}^{(s)} \cdot \mathbf{u}^{(n-1)}(z_\ell)\right)(\mathbf{I} - \Delta_n z_\ell \mathbf{A})^{-1}\mathbf{1}$$

in the following way.

Algorithm 28 (Runge-Kutta gCQ with contour quadrature)

- ***Initialization.*** *Generate $\mathcal{K}_{-\rho}(z_\ell)$ for all contour quadrature nodes z_ℓ, $\ell = 1, 2, \ldots, N_Q$. Compute $\widetilde{\boldsymbol{\phi}}^{(1)}$ from*

$$K_{-\rho}\left((\Delta_1 \mathbf{A})^{-1}\right)\widetilde{\boldsymbol{\phi}}^{(1)} = \partial_t^\rho \mathbf{g}^{(1)}. \tag{88}$$

- ***For*** $n = 2, \ldots, N$

  1. ***Runge–Kutta step.*** *Perform a step of the Runge–Kutta method applied to* (8b) *and compute*

  $$\mathbf{u}^{(n-1)}(z_\ell) = (\mathbf{I} - \Delta_{n-1} z_\ell \mathbf{A})^{-1}\left((\mathbf{1} \otimes \mathbf{e}_s)\,\mathbf{u}^{(n-2)}(z_\ell) + \Delta_{n-1}\mathbf{A}\widetilde{\boldsymbol{\phi}}^{(n-1)}\right)$$

  for all contour quadrature nodes: $z = z_\ell$, $\ell = 1, \ldots, N_Q$.

  2. ***Generate linear system.*** *If Δ_n is a new time step, then generate $K_{-\rho}\left((\Delta_n \mathbf{A})^{-1}\right)$. Otherwise this operator was already generated in a previous step. Update the right-hand side*

  $$\mathbf{r}^{(n)} = \mathbf{r}^{(n)}\left(\mathbf{u}^{(n-1)}\right) := \partial_t^\rho \mathbf{g}^{(n)} - \sum_{\ell=1}^{N_Q} w_\ell K_{-\rho}(z_\ell)\left(\mathbf{e}^{(s)} \cdot \mathbf{u}^{(n-1)}(z_\ell)\right)(\mathbf{I} - \Delta_n z_\ell \mathbf{A})^{-1}\mathbf{1}.$$

  3. ***Linear Solve.*** *Solve the linear system*

  $$\mathcal{K}_{-\rho}\left((\Delta_n \mathbf{A})^{-1}\right)\widetilde{\boldsymbol{\phi}}^{(n)} = \mathbf{r}^{(n)}.$$

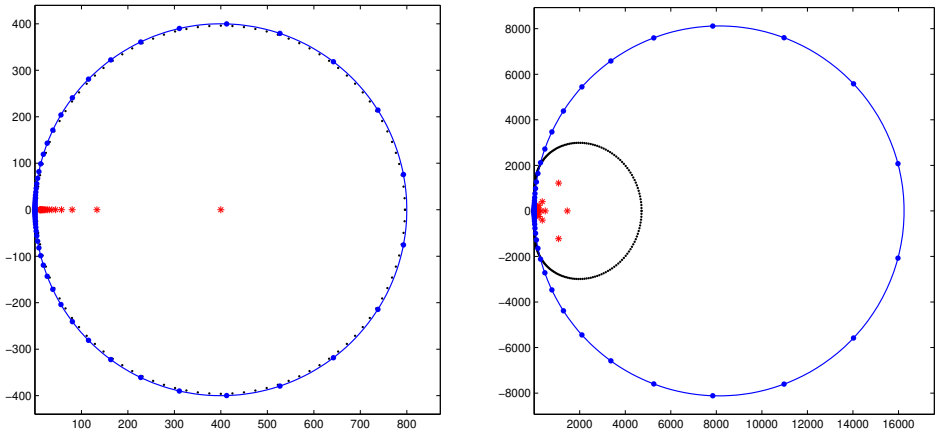

Figure 1: Poles of the integrand in (22), integration contour and curve $|R(\Delta_{\min} z)| = 1$ for 20 steps quadratically graded towards the origin. *Left:* For implicit Euler method. *Right:* For RadauIIA5

For gCQ based on the implicit Euler method the quadrature problem has been fully solved in [9] and several experiments are reported in [10]. The contour of choice in this case is the circle centered at $\Delta_{\min}^{-1}$ with radius $\Delta_{\min}^{-1}$, which coincides with the boundary of the region $|R(\Delta_{\min} z)| = 1$. The parameterization of this circle uses Jacobi elliptic functions in order to optimally exploit the analyticity domain of the integrand in (22), whose poles are located in the real segment $[\Delta^{-1}, \Delta_{\min}^{-1}]$.

For higher order Runge–Kutta methods the poles of the integrand in (22) are typically located in a sector around the positive real axis and the boundary of the stability region $|R(\Delta_{\min} z)| = 1$ is more complicated than a circle. In Figure 1 we show the location of the poles, the curve $|R(\Delta_{\min} z)| = 1$ and our contour of choice for the grid

$$t_j = \left(\frac{j}{20}\right)^2, \quad j = 1, \ldots, 20,$$

both for implicit Euler and RadauIIA5. In both cases we choose a circle as the integration contour but in the case of RadauIIA5 the radius is much larger, namely $M = 5 \max(|\lambda|)/\Delta_{\min}$ for $\lambda \in \sigma(\mathbf{A})$. This implies that the boundary of the contour becomes more vertical at $z = 0$ and thus avoids invading too much into the region $|R(\Delta_{\min} z)| > 1$ close to the origin. For this contour the number of quadrature nodes needed to produce the error plot in Figure 2 was $N_Q = 3N \log^2(N)$. The optimization of the integration contour and a rigorous error and complexity analysis are the subject of ongoing research.

In order to illustrate the performance of high order Runge–Kutta gCQ in comparison with the original CQ, with uniform steps, we consider the following

one-dimensional example: Find ϕ such that $K(\partial_t)\phi = g$ with

$$K(z) = \frac{1 - \mathrm{e}^{-2z}}{2z} \quad \text{and} \quad g(t) = t^{5/2}\mathrm{e}^{-t}. \tag{89}$$

The exact solution to this problem is computed in [17] and is given by

$$\phi(t) = 2 \sum_{k=0}^{\lfloor t/2 \rfloor} g'(t - 2k). \tag{90}$$

We approximate $\phi(t)$ for $t \in [0, 1]$ by applying Algorithm 28 for with RadauIIA5 and $\rho = 0$. Then we have $\mu = 1$ in Assumption 1, $p = 5$ and $q = 3$. The right-hand side g satisfies $g^{(\ell)}(0) = 0$ for $\ell = 0, 1, 2$ and is not three times differentiable at $t = 0$. This lack of regularity suggests to use a time grid which is algebraically graded towards the origin. We heuristically choose a quadratically graded mesh with points

$$\Theta = (t_j)_{j=1}^N \quad \text{with} \quad t_j = \left(\frac{j}{N}\right)^\alpha$$

and $\alpha = 2$. In this case it is $\Delta = N^{-1}$ and $\Delta_{\min} = N^{-2}$. For a comparison with uniform steps we set $\alpha = 1$. Figure 2 shows that the convergence rate is $\mathcal{O}(\Delta^3)$ for the graded mesh and about $\mathcal{O}(\Delta^{1.6})$ for the uniform mesh. For this example, we have $\mu = 1$ and thus the minimal integer $\nu > \mu + 1$ is $\nu = 3$. For $\rho = 0$, with $\mu - \rho = 1 > -1$, Theorem 14 then predicts a convergence rate like $\mathcal{O}(\Delta^{3+1-3}) = \mathcal{O}(\Delta)$. The theoretical estimate provided by this Theorem is of order 3 or higher only for $\rho \geq 2$. More precisely it is $\mathcal{O}\left(\Delta^3\right)$ for $\rho = 2$ and $\mathcal{O}(\Delta^5)$ for $\rho = 3$. We believe this is due to a limitation of our theory which does not allow in principle to choose a fractional value of ν. In the limit (not allowed) case $\nu = 2$, the theoretical estimate yields actually an estimate like $\mathcal{O}(\Delta^2)$. However our numerical result for $\rho = 0$ is better and actually coincides with the theory for uniform steps developed in [1]. It is an open problem whether there exist examples where a bigger value of ρ is necessary for variable steps than for uniform steps or whether our theory yields a suboptimal estimate in terms of this parameter.

References

[1] L. Banjai and C. Lubich. An error analysis of Runge-Kutta convolution quadrature. *BIT*, 51(3):483–496, 2011.

[2] L. Banjai, C. Lubich, and J. M. Melenk. Runge-Kutta convolution quadrature for operators arising in wave propagation. *Numer. Math.*, 119(1):1–20, 2011.

[3] L. Banjai and S. Sauter. Rapid solution of the wave equation in unbounded domains. *SIAM Journal on Numerical Analysis*, 47:227–249, 2008.

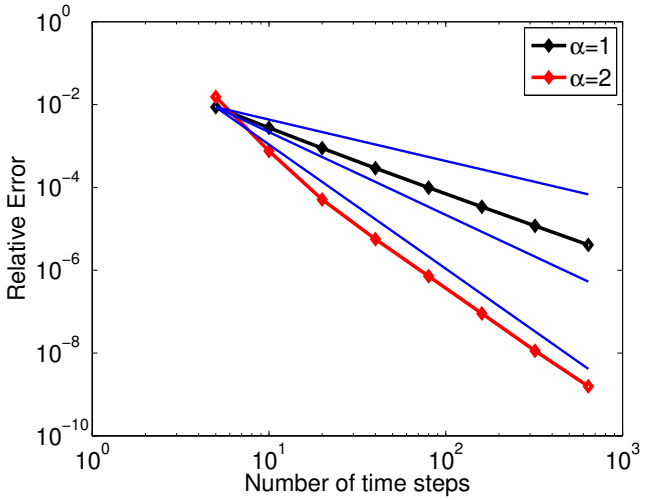

Figure 2: Error with respect to the number of steps for g in (89). The straight lines indicate slopes 1, 2 and 3, respectively.

[4] C. de Boor. Divided differences. *Surv. Approx. Theory*, 1:46–69, 2005.

[5] S. Falletta, G. Monegato, and L. Scuderi. A space-time BIE method for nonhomogeneous exterior wave equation problems. The Dirichlet case. *IMA J. Numer. Anal.*, 32(1):202–226, 2012.

[6] W. Greub. *Linear Algebra*. Springer-Verlag, New York, fourth edition, 1975.

[7] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer, 2012.

[8] E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2010. Stiff and differential-algebraic problems, Second revised edition, paperback.

[9] M. Lopez-Fernandez and S. A. Sauter. Fast and Stable Contour Integration for High Order Divided Differences via Elliptic Functions. Technical Report 08-2012, Institut für Mathematik, Univ. Zürich, 2012. to appear in MathComp.

[10] M. Lopez-Fernandez and S. A. Sauter. A Generalized Convolution Quadrature with Variable Time Stepping. Part II: Algorithms and Numerical Results. Technical Report 09-2012, Institut für Mathematik, Univ. Zürich, 2012.

[11] M. Lopez-Fernandez and S. A. Sauter. Generalized Convolution Quadrature with Variable Time Stepping. *IMA J. Numer. Anal.*, 33(4):1156–1175, 2013.

[12] C. Lubich. Convolution Quadrature and Discretized Operational Calculus I. *Numerische Mathematik*, 52:129–145, 1988.

[13] C. Lubich. Convolution Quadrature and Discretized Operational Calculus II. *Numerische Mathematik*, 52:413–425, 1988.

[14] C. Lubich. On the multistep time discretization of linear initial-boundary value problems and their boundary integral equations. *Numerische Mathematik*, 67(3):365–389, 1994.

[15] C. Lubich. Convolution quadrature revisited. *BIT Numerical Mathematics*, 44:503–514, 2004.

[16] C. Lubich and A. Ostermann. Runge-Kutta methods for parabolic equations and convolution quadrature. *Math. Comp.*, 60(201):105–131, 1993.

[17] S. Sauter and A. Veit. A Galerkin Method for Retarded Boundary Integral Equations with Smooth and Compactly Supported Temporal Basis Functions. Part II: Implementation and reference solutions. *Numer. Math.*, 2012. electronic.